The Subjective Feelings of Comprehension and Remembering Accompanying Text Learning On-Screen

Rakefet Ackerman
University of Haifa
rakefet@research.haifa.ac.il

Abstract

Learners still prefer studying text in print despite numerous attempts to improve subjective experience that accompanying text learning on screen. Using a metacognitive approach, the differences between on-screen and on-paper learning were analyzed by differentiating comprehension and memorizing aspects of learning. Participants studied expository texts of 1000-1200 words in one of the two media. They predicted their performance for comprehension questions and for details memorizing before being tested. Under fixed study time (Experiment 1) performance did not differ between the two media, but when free control over study time was allowed (Experiment 2), performance was lower on screen than on paper. Comprehension predictions were less accurate than memorizing predictions overall, and especially for on-screen learning. The findings lead to rejection of ergonomic factors as the main cause for screen inferiority and point instead to monitoring of comprehension as a potential source for the subjective and objective difference between on-screen and on-paper learning.

Keywords: Metacognition, meta-comprehension, e-learning, text learning.

Introduction

The extensive use of computers and the development in display technology and software design has not changed people's preference for print over digitally presented text when thorough study is required (e.g., Ackerman & Goldsmith, 2008; Annand, 2008; Buzzetto-More, Sweat-Guy, & Elobaid, 2007; Spencer, 2006). The present study investigated the basis for subjective screen inferiority using a metacognitive approach. This approach emphasizes that only when students assess accurately the state of their ongoing knowledge acquisition can they direct learning by allocating study time effectively and selecting appropriate study activity (Brown, Smiley, & Lawton, 1978; Pressley, 2002; Son & Schwartz, 2002). A widely accepted set of models for control over study time are describes as Discrepancy Reduction Models (Butler & Winne, 1995; Dunlosky & Hertzog, 1997; Thiede & Dunlosky, 1999). This framework describes the relationship between monitoring, judgment of current state of knowledge, and control, study activities directed to achieve learner's goals (Nelson & Narens, 1990). According to the Discrepancy Reduction Models, the judged level of knowledge is compared during study with a preset target-level. Study continues until this target level has apparently been reached. Indeed, control over study was found to be based on subjective knowledge rather than on the objective state of learning (Metcalfe & Finn, 2008; Thiede, Anderson, & Therriault, 2003).

Monitoring accuracy measures the correspondence between judgment and knowledge level. Two measures have been commonly used: *calibration* and *resolution* (e.g., Nelson & Dunlosky, 1991). In the text-learning context, *calibration* measures the realism of judgment—the disparity

between prediction of performance (POP) and test score. If generally overconfident, then at some point during study the learner will believe that the target level has been reached whereas in fact it has not. Alternatively, underconfident learner may continue studying materials that are already adequately mastered, while this study time could be invested more effectively in other study materials. Accurate judgment of learning is therefore expected to yield the best results. *Resolution* measures judgment differentiation between better known and less known materials. Perfect resolution is achieved if every text judged by higher POP indeed got higher test score than every text with lower POP. This differentiation is important for efficient choice of materials for study or for restudy (Metcalfe & Finn, 2008; Thiede & Dunlosky, 1999) and affects study regulation efficiency and test performance (Thiede et al., 2003).

Ackerman and Goldsmith (2008) examined the relationship between monitoring accuracy and control over study time for on-screen learning (OSL) and for on-paper learning (OPL) of expository texts. They found that performance was equivalent for OSL and OPL when control over study time was prevented by fixing the allotted study time per text, but was lower for OSL than for OPL when free control over study time was permitted. If ergonomic factors were responsible for the lower performance on screen, then it would be expected to affect the result under both study schedules. The fact that performance inferiority of OSL was revealed just when free control over study time was allowed, pointed to metacognitive factors as responsible for the found difference. Moreover, on-paper learners, but not on-screen learners, benefited from self-regulation and additional invested study time, leading to higher scores than in the prefixed study time condition. Ackerman and Goldsmith's methodology provided a *Metacognitive Learning Regulation Profile* of each medium. Measures of monitoring accuracy showed no resolution difference between the two media, but a consistent calibration difference. OSL participants were overconfident in their POP in all study conditions, whereas OPL participants were almost accurate. The present study asks: What aspect of the metacognitive monitoring generated the difference between OSL and OPL: monitoring of higher level comprehension, or monitoring of details memorizing?

Most metacognitive studies of text-learning monitoring have been used overall POP, although some of them included higher inference questions and details memorizing questions in their knowledge test (e.g., Dunlosky & Rawson, 2005; Thiede et al., 2003). Under this procedure, participants determined whether to emphasize higher level inferences or memorizing in their judgment. However, for the purpose of reading comprehension, monitoring should take place in several levels in parallel for achieving representation of words, sentences, and higher-level propositions (Kintsch, 1998). Rawson, Dunlosky, and McDonald (2002) examined the difference between comprehension judgment and overall prediction of performance at test time. They found that comprehension judgments were higher than POPs and that only POPs were affected by manipulation of expected test delay, indicating that at least some role was given to retention over time within overall POP. Maki (1995) compared overall POP to performance in higher level inference questions and in questions regarding specific important sentences in texts presented on screen. She found that POPs were more predictive of performance in higher order questions than in questions tapping memory for specific sentences in the text. This finding suggests an emphasis of participants' judgments toward higher order understanding. I suggest resolving the doubt by asking participants to provide two POPs, one for comprehension and one for memorizing, and examining their monitoring accuracy relative to their performance on respective test questions.

Experiment 1

The purpose of Experiment 1 was to expose the differentiation between monitoring comprehension and monitoring memorizing in the two media before control over study time took effect on final POP. Undergraduate students (N = 34) studied six exploratory texts of about 1000 words each, taken from web sites intended for reading on screen. Half of them studied on screen and half on paper. They stopped studying after prefixed and short time period (7 minutes). POP was provided separately for comprehension of high order ideas and for details memorizing, taking into account the limited time allowed for study. The exam included five questions regarding inference of higher order ideas and five questions regarding details in the text, all intermixed.

Looking at the media difference, no performance difference was found between the media in both comprehension and memorizing question types (See Figure 1, Panel A).

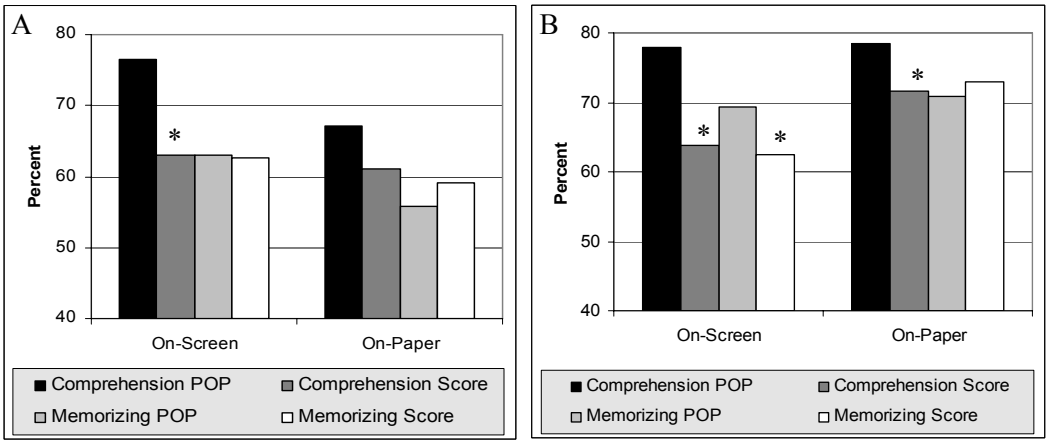

Figure 1. Comprehension and memorizing POP and the respective scores in Experiment 1 with fixed study time (Panel A) and in Experiment 2 with self-paced study schedule (Panel B)

POP comparison by a two-way Analysis of Variance (ANOVA) of Question Type × Media yielded a main effect of question type $F(1,32) = 93.36$, $p < .0001$ and a main effect of media $F(1,32) = 5.56$, $p < .05$. Thus, regardless of the media, participants differentiated between comprehension and memorizing monitoring and provided higher POPs for comprehension than for memorizing. Calibration, as measured by the difference between POP and test score, was accurate for OPL whereas significant overconfidence was found for OSL $t(16) = 2.48$, $p < .05$. This overconfidence stemmed from a significant overconfidence in OSL comprehension judgment $t(16) = 4.20$, $p < .01$. Thus, while comprehension judgment tended to be biased in both media, this bias was especially pronounced on screen.

Resolution did not differ between OSL and OPL for either of the POP types when measured by within-participant gamma correlation (Nelson, 1984) as well as by Confidence Accuracy Quotient (CAQ) score[1], that fits better small number of judgment-score pairs paradigms (see Lundeberg, Fox, & PunCochar, 1994; Shaughnessy, 1979).

[1] $$CAQ = \frac{M_{higher} - M_{lower}}{SD}$$

M_{higher} is the mean test score of the three texts with the higher POPs. M_{lower} is the mean score of the three texts for which lower POP were given. SD is the standard deviation of test scores for each participant. A CAQ score of zero denotes someone who has no resolution of monitoring; positive CAQ score indicates higher judgment when higher rather than lower scores were achieved; negative CAQs denote resolution in the wrong direction (i.e., higher score when POP was lower than when it was higher).

Experiment 2

In Experiment 2 participants (N = 74) had full control over their study time per text, within a global time limit (90 minutes). The purpose was to examine whether there was a difference between OSL and OPL in the ability of participants to take advantage of the opportunity to regulate their study and examine the role of the two levels of monitoring in the achieved performance level.

In contrast to the performance similarity between the two media in Experiment 1, here—under self-paced study time—the scores were lower for OSL than for OPL, overall and in separation into the two questions types. A two-way ANOVA of Question Type × Media on test score revealed a main effect of media $F(1,72) = 11.17$, $p = .001$ with no significant interaction. Somewhat shorter study time was invested for studying each text by OSL participants relative to OPL participants $t(72) = 1.71$, $p = .09$.

Overall, as in Experiment 1, participants provided higher comprehension POPs than memorizing POPs $F(1,73) = 99.37$, $p < .0001$. Monitoring accuracy of the two question types is the focus of the present study. Turning first to examine calibration, a two-way ANOVA as before on overconfidence revealed a main effect of question type $F(1,72) = 31.34$, $p < .0001$ and a main effect of media $F(1,72) = 6.78$, $p = .01$. Thus, comprehension POP suffered larger overconfidence than memorizing POP and overall larger overconfidence was found for OSL than for OPL. See Figure 1, Panel B.

When seeking to explain overconfidence in comprehension, one possibility is that participants reflect an evaluation of general ability to comprehend, whereas memorizing judgment is more materials-related. In this case, we would expect low variability in comprehension relative to memorizing judgment. Indeed, a two-way ANOVA as before on within-participant standard deviation (SD) revealed lower variability of comprehension POP than of memorizing POP $F(1,72) = 20.71$, $p < .0001$, with no interaction with the media. However, the absolute means difference was small (8.3 for comprehension; 10.1 for memorizing) and both SDs were significantly different from zero $p < .0001$. This finding suggests that both POPs were not based solely on overall preliminary evaluation of participants' learning ability.

As explained above, the second important aspect of monitoring accuracy is resolution. Overall resolution did not differ between the two media, as in Experiment 1. Separation into question types exposed no difference between the media in memorizing resolution. However, virtually no comprehension resolution was found for OSL whereas significantly higher comprehension resolution was found for OPL $t(72) = 2.21$, $p < .05$. See Table 1.

Table 1. POP prediction of actual score, in the target question type and in the other question type

Media	Overall resolution	Comprehension resolution		Memorizing resolution	
		Target type: Comprehension	Other type: Memorizing	Target type: Comprehension	Other type: Memorizing
On-Screen	.39[a]	.02	.16	.49[a]	.21
On-Paper	.36[a]	.43[a]	.16	.56[a]	.14

[a] Significiantly different from zero

In order to further examine the differentiation between the two judgment types, I compared the predictability of memorizing POP for memorizing score, its target question type, relative to the predictability of comprehension scores, the other question type. The same was also done for

examining comprehension POP predictibility. It can be seen from Table 1 that in most of the cases the predictibility of POP for its target question type was significant and higher than the prediction of the other question type. This finding supports once again the differentiation participants do between the two POP types. However, resolution of comprehension POP for OSL was an exception as it was very weak predictor for both types of question.

Summary

OSL inferiority relative to OPL was indeed found in this study under natural study conditions. This finding supports the common subjective feeling accompanying OSL. However, the similar performance level under fixed study-time schedule directs us to reject ergonomic factors as responsible for this finding. Alternatively, this similarity emphasizes the role of metacognitive monitoring accuracy in exposing the source for the difference between OSL and OPL. The separation of comprehension and memorizing judgments was proposed in order to attract participants' attention to the difference between the monitored knowledge levels and to generate appropriate test expectation (Thiede, Wiley, & Griffin, 2007). This procedure revealed that learners differentiated between the two judgment types and exposed differences between them in both aspects of monitoring accuracy: calibration and resolution. These findings are important beyond the issue of media comparison.

The fairly accurate monitoring under OPL showed that participants could have good basis for their study regulation (cue diagnosticity; Baker & Dunlosky, 2007). This finding emphasizes OSL inferiority in monitoring accuracy. The findings suggest that comprehension monitoring originated the main monitoring fault in OSL, whereas better monitoring accuracy achieved for memorizing.

An important question is raised: Why do people feel less secure when studying on-screen, but their per-text judgment shows consistent overconfidence? It seems that OSL monitoring is not directly indicative of knowledge gain. If this weak relationship between monitoring and performance can be monitored at a higher level, it might be a justifiable basis for the overall reluctance that accompanies screen learning. The study conditions generated here, that people study the same materials but show consistent differential monitoring quality, is unique and points to this speculation. Moreover, it calls for applying the Metacognitive Learning Regulation Profile for analyses of other study conditions that accompany differential subjective feelings for the same study tasks.

Acknowledgement

Support for the preparation of this article by The Inter-University Center For E-Learning (IUCEL - MEITAL) is gratefully acknowledged.

References

Ackerman, R., & Goldsmith, M. (2008). Learning directly from screen? Oh-no, I must print it! Metacognitive analysis of digitally presented text learning. In Y. Eshet, A. Caspi, & N. Geri (Eds.), *Learning in the technological era* (Vol. 3, pp. 1-7). Raanana, Israel: Open University of Israel.

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.

Baker, J. M. C., & Dunlosky, J. (2007). The relationship between cue diagnosticity and judgment accuracy for judgments of text learning. Paper presented at *the 48th Annual Meeting of the Psychonomic Society*, Long Beach, CA.

Brown, A. L., Smiley, S. S., & Lawton, S. Q. C. (1978). The effects of experience on the selection of suitable retrieval cues for studying texts. *Child Development, 49*, 829-835.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245.

Buzzetto-More, N., Sweat-Guy, R., & Elobaid, M. (2007). Reading in a digital age: e-books: are students ready for this learning object. *Interdisciplinary Journal of Knowledge and Learning Objects, 3*, 239-250.

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 52*, 178-186.

Dunlosky, J., & Rawson, K. A. (2005). Why does rereading improve metacomprehension accuracy? Evaluating the Levels-of-disruption hypothesis for the rereading effect. *Discourse Processes, 40l*, 37-55.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge.

Lundeberg, M. A., Fox, P. W., & Punccohar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology, 86*, 114-121.

Maki, R. H. (1995). Accuracy of metacomprehension judgments for questions of varying importance levels. *American Journal of Psychology, 108*, 327-344.

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

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109-133.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science, 2*, 267-270.

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, pp. 125-173). San Diego, CA: Academic Press.

Pressley, M. (2002). Metacognition and self-regulated comprehension. *What research has to say about reading instruction, 3*, 291-309.

Rawson, K. A., Dunlosky, J., & McDonald, S. L. (2002). Influences of metamemory on performance predictions for text. *The Quarterly Journal of Experimental Psychology Section A, 55*, 505-524.

Shaughnessy, J. J. (1979). Confidence-judgment accuracy as a predictor of test performance. *Journal of Research in Personality, 13*, 505-514.

Son, L. K., & Schwartz, B. L. (2002). The relation between metacognitive monitoring and control. In T. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 15-38). Cambridge: Cambridge University Press.

Spencer, C. (2006). Research on learners' preferences for reading from a printed text or from a computer screen. *Journal of Distance Education, 21*, 33-50.

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

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: an analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024-1037.

Thiede, K. W., Wiley, J., & Griffin, T. T. (2007). Does test expectancy influence metacomprehension accuracy? Paper presented at the *48th Annual Meeting of the Psychonomic Society*, Long Beach, CA.