Metacognitive Regulation of Text Learning: On Screen Versus on Paper

Rakefet Ackerman
Technion–Israel Institute of Technology

Morris Goldsmith
University of Haifa

Despite immense technological advances, learners still prefer studying text from printed hardcopy rather than from computer screens. Subjective and objective differences between on-screen and on-paper learning were examined in terms of a set of cognitive and metacognitive components, comprising a *Metacognitive Learning Regulation Profile* (MLRP) for each study media. Participants studied expository texts of 1000–1200 words in one of the two media and for each text they provided metacognitive prediction-of-performance judgments with respect to a subsequent multiple-choice test. Under fixed study time (Experiment 1), test performance did not differ between the two media, but when study time was self-regulated (Experiment 2) worse performance was observed on screen than on paper. The results suggest that the primary differences between the two study media are not cognitive but rather *metacognitive*—less accurate prediction of performance and more erratic study-time regulation on screen than on paper. More generally, this study highlights the contribution of metacognitive regulatory processes to learning and demonstrates the potential of the MLRP methodology for revealing the source of subjective and objective differences in study performance among study conditions.

Keywords: metacognition, metacomprehension, monitoring and control, text learning, self-regulated learning, computer-based learning

Adult readers of today have been using computers extensively for many years. Nevertheless, when one needs to study a text thoroughly, there is still a strong preference to print out digital text rather than study it directly from the computer screen (Buzzetto-More, Sweat-Guy, & Elobaid, 2007; Dilevko & Gottlieb, 2002; Spencer, 2006). One might assume that this reluctance is a matter of experience. However, even highly experienced computer users still prefer print, as Buxton (2008), principal researcher at Microsoft, admits: "I can't stand reading stuff on my computer" (p. 8).

Objective and subjective learning differences between paper and screen learning have been examined and discussed for some time. Dillon, McKnight, and Richardson (1988), for example, pointed to differences by which reading texts on screen is slower, less accurate, more fatiguing, accompanied by reduced comprehension, and subjectively less effective than reading from paper. Additional studies examined the effects of factors such as technical characteristics of displays, annotation while reading, navigation ease, and spatial layout on reader preferences and performance (e.g., Dillon, Richardson, & McKnight, 1990; O'Hara & Sellen, 1997; Richardson, Dillon, & McKnight, 1989; see Dillon, 1992). Of course, technology has improved dramatically in the 20 years or so since those results were obtained. Hence, one might question whether such findings are still relevant. Recent findings, however, indicate a dislike of on-screen reading even among young adults studying with current state-of-the-art displays (e.g., Annand, 2008; Eshet-Alkalai & Geri, 2007; Rogers, 2006; Shaikh, 2004; Spencer, 2006). From the point of view of hardware engineers and software designers, this persistent reluctance to read "serious" texts on screen indicates that the large effort invested in improving reading from computer screens (e.g., Dillon, 1994; Muter, 1996) has not yet achieved its goals (Dillon, 2002; Garland & Noyes, 2004; Rogers, 2006; Sellen & Harper, 2002).

The general finding of both objective and subjective difficulties related to on-screen learning makes this an ideal topic for examination from a metacognitive perspective. A great deal of research on metacognition and learning has revealed the crucial role that subjective experience plays in guiding and regulating the learning process: in the choice of study strategy, in the allocation and prioritization of study time, in deciding when one has sufficiently mastered the material, and so forth (Baker, 1985; Bielaczyc, Pirolli, & Brown, 1995; Bjork, 1994; Brown, Smiley, & Lawton, 1978; Hacker, 1998; Schunk & Zimmerman, 1994; Son, 2007). Adopting such a perspective, the present study examined whether subjective differences between on-screen learning (OSL) and on-paper learning (OPL) might in fact underlie, rather than merely reflect, objective differences in learning performance.

Assessing Metacognitive Regulation of Text Learning

In general, metacognitive theories of learning address the interplay between objective and subjective aspects of the learning process. Building on concepts and measures from leading metacognitive theories of the control of study time, in the present study objective (cognitive) and subjective (metacognitive) aspects of expository text learning were assessed and compared using a new *Metacognitive Learning Regulation Profile* (MLRP) methodology, which provides a multicomponential assessment of learning processes under specific conditions. The comparison between the MLRPs of OSL and OPL allowed heretofore hidden differences in underlying components of the learning process between the two media to be revealed. Table 1 presents the MLRP components that were examined. These will now be explained.

According to the highly influential *discrepancy reduction model* (Butler & Winne, 1995; Dunlosky & Thiede, 1998; Nelson & Narens, 1990; for a graphic depiction, see Winne & Hadwin, 1998), people begin their study activity by setting a target level of learning (Le Ny, Denhiere, & Le Taillanter, 1972). Allocation of study time is then guided by an ongoing subjective assessment of knowledge level and comparison to the preset target level: when the subjective knowledge level is satisfactory, that is, when the target level is reached for a particular item, the learner terminates the study of that item and moves on to another. Hypothesized study curves based on the model are depicted in Figure 1 and will be explained shortly.

Basic measures of study regulation that stem from the discrepancy reduction model are as follows: *Prediction of performance* (*POP*; Maki, 1998b)*, study time*, and *test performance* (components 2, 7, and 8 in Table 1). POP reflects learners' ongoing monitoring and final subjective assessment of their level of knowledge, tapped by having them predict their future test performance after studying each text (Maki & Serra, 1992; Rawson, Dunlosky, & McDonald, 2002).[1] Study time is an objective measure that is assumed to reflect the metacognitive control decision to continue or to terminate study, based on the ongoing monitoring of knowledge level. Test performance is, of course, the ultimate objective measure of the learner's success. The relationships among these three measures produce additional measures that are of interest. *Encoding efficiency* can be examined in terms of the amount of information stored (and retained) into memory during a fixed amount of study time, that is, when learners have no control over study time (component 1 in Table 1). This measure yields information about the "objective" efficiency of learning, controlling for possible differences in the effectiveness of the subjective knowledge monitoring and study-time control decisions that contribute to self-regulated study.

The MLRP components that reflect the accuracy of the *metacognitive monitoring* are *calibration bias* and *resolution* (components 3 and 4 in Table 1). *Calibration bias*, or absolute monitoring accuracy, is calculated as the mean signed deviation between POP and test score (e.g., Metcalfe, 1998). Based on the discrepancy reduction model, study should be terminated when POP reaches the target level of mastery (see Figure 1, stopping points A and B). Hence, underconfidence, an overly low subjective assessment of knowledge, will lengthen the study time unnecessarily, wasting time that could perhaps be invested more effectively in other materials. By contrast, overconfidence, which is the more common

situation (Metcalfe, 1998), will lead to premature study termination and a lower than desired level of performance (e.g., Glenberg, Wilkinson, & Epstein, 1982; Pressley & Ghatala, 1988). See Figure 1, stopping point A.

Resolution, or relative monitoring accuracy, indexes the extent to which POP discriminates between learned and unlearned information (Glenberg, Sanocki, Epstein, & Morris, 1987; Lundeberg, Fox, & Punćochaŕ, 1994; Maki & Serra, 1992; Rawson, Dunlosky, & Thiede, 2000; Thiede, Anderson, & Therriault, 2003). Resolution is maximized when all of the better-learned texts are assigned a higher predicted performance than all of the lesser-learned texts. This discrimination can then be used by learners to select the most appropriate material for extra study (Metcalfe & Finn, 2008; Thiede & Dunlosky, 1999). Relatively low levels of monitoring resolution have been found in text learning (see Maki, 1998b for a review), but resolution was improved by some techniques such as by using immediate rather than delayed prediction (Maki, 1998a) and by asking for predictions that relate to performance across multiple test questions (Weaver, 1990).

Moving on to measures of *metacognitive control*, the discrepancy reduction model entails two factors that relate to the efficiency of control over study time: *control criterion* and *control sensitivity* (components 5 and 6 in Table 1). *Control criterion* ("norm of study"; Le Ny et al., 1972) refers to the target level of knowledge implicitly set by the learner to guide his or her study-time control decisions. According to the discrepancy reduction model, the higher the criterion that is set for a particular text, the better that text will be learned (cf. Nelson & Leonesio, 1988), but this will generally require a greater investment of study time, which may leave less time for other material. Thus, the setting of the control criterion should be strategic in nature and influenced by motivational and situational factors (cf. similar ideas with regard to controlling the information that is reported from memory; e.g., Ackerman & Goldsmith, 2008a; Goldsmith & Koriat, 2008). Support for this idea comes from studies showing that increasing the rewards for correct answers to particular items increases the amount of time devoted to studying those items (Dunlosky & Thiede, 1998), as does emphasizing accuracy over speed of learning (Lockl & Schneider, 2004; Nelson & Leonesio, 1988), as does setting the "passing grade" for an upcoming test at 90% rather than at 25% correct (LaPorte & Nath, 1976). With regard to potential differences between OSL and OPL, if different target levels of learning are being set for each media, perhaps because of different levels of motivation or subjective comfort, one would then expect to find correspondingly different subjective and objective levels of performance between the two media.

A final component of the MLRP is *control sensitivity*: the extent to which the learner's control decisions are in fact sensitive to his or her subjective monitoring. Essentially, this factor refers to the tightness of the relationship between the control operation and the monitoring judgment on which it is assumed to be based. In work guided by their model of the strategic regulation of memory

---

[1] In the metacognitive literature relating to the memorization of word lists, this prediction is commonly termed *Judgment of Learning* (JOL). Because in the present article our specific focus is on the study of texts, we adopt the term POP to indicate the more complex, multilevel nature of metacognitive assessments relating to the mastery of textual material.

Table 1
Summary Comparison of Metacognitive Learning Regulation Profiles (MLRP) for On-Screen Learning (OSL) Versus On-Paper Learning (OPL), Across the Experiments

MLRP Component	Component source	Qualitative comparison result
Cognitive (encoding/storage)		
1. Encoding efficiency	Experiment 1	OSL = OPL
Metacognitive monitoring		
2. Prediction of performance (POP)	Experiment 1	OSL > OPL
	Experiment 2 First online POP	OSL > OPL
	Experiment 2 Terminal POP	OSL = OPL
3. Calibration bias	Experiment 1	OSL > OPL
	Experiment 2	OSL > OPL
4. Resolution	Experiment 1	OSL = OPL
	Experiment 2	OSL = OPL
Metacognitive control		
5. Control criterion	Experiment 2 Online POP	OSL = OPL
6. Control sensitivity	Experiment 2 Online POP	OSL < OPL
Performance		
7. Self-regulated study time	Experiment 2 Terminal POP	OSL < OPL
8. Self-regulated test performance	Experiment 2	OSL < OPL

reporting, Koriat and Goldsmith (1996; for a review, see Goldsmith & Koriat, 2008) defined control sensitivity as the correlation between subjective confidence in the correctness of a potential answer (the monitoring output) and the decision whether to report it or respond "don't know" (the control decision). This correlation was found to reach near-ceiling levels with healthy undergraduate participants (Koriat & Goldsmith, 1996). In other studies with special populations, however, this very high level of control sensitivity was reduced, offering insights into the nature of the cognitive and metacognitive deficits ensuing from old age (Pansky, Koriat, Goldsmith, & Pearlman-Avnion, 2009) and mental illness (Danion, Gokalsing, Robert, Massin-Krauss, & Bacon, 2001; Koren et al., 2004).

In the context of study regulation, control sensitivity can be examined in terms of the consistency of the relationship between POP [(or judgment of learning (JOL)] and the control of study time. In general, a strong relationship has been found between JOL

Figure 1. Illustrative objective and subjective hypothetical learning curves, based on the discrepancy reduction model, comparing conditions of perfect calibration and overconfidence. Differences in knowledge level at three different study termination points are also indicated: A – termination with overconfidence, B – termination with perfect calibration, and C – termination after a short, fixed study time (as in Experiment 1, here).

and the allocation of study time (e.g., Mazzoni, Cornoldi, & Marchitelli, 1990), indicating a high level of control sensitivity. In this context too, however, differences in control sensitivity may underlie population differences in the effectiveness of study regulation. For example, Lockl and Schneider (2004) found that although children as young as 6 years old can state that pairs of related words are easier to learn than unrelated word pairs (Dufresne & Kobasigawa, 1989), nine-year-old children, but not seven-year-old children, allocated study time in accordance with item difficulty, suggesting lower control sensitivity in the younger group (see also Koriat, Ackerman, Lockl, & Schneider, 2009). Similarly, results from Dunlosky and Connor (1997) suggest that older adults do not use online monitoring to allocate study time to the same degree as younger adults do and that these allocation differences contribute to age deficits in recall.

An underlying assumption of the discrepancy reduction model is that control sensitivity is high—learners continue studying as long as POP is below the criterion (target) level and stop studying as soon as the criterion is reached (see Figure 1, stopping points A and B). Thus, the measure of control sensitivity implied by this model is the strength of the relationship between the ongoing POP level during the study and the decision to continue or to stop studying.

A somewhat different conception of control sensitivity is implied by an alternative model for control over study time, proposed by Metcalfe and Kornell (2003, 2005). According to the *region of proximal learning model*, people base their decision to stop studying on the perceived rate at which learning progresses, rather than on a comparison of the absolute judgment level to a predefined control criterion. When the learners perceive that they are gaining knowledge at a rapid rate, they continue. When they feel that they are no longer taking in information, they stop studying a particular item and switch to another. Thus, monitoring of knowledge gaining rate is expected to be the basis for appropriate control decisions (Son & Metcalfe, 2000). The region of proximal learning model was used mainly to explain the finding that people allocate more study time to intermediate-difficulty material than to the most difficult material, as would be predicted by the discrepancy reduc-
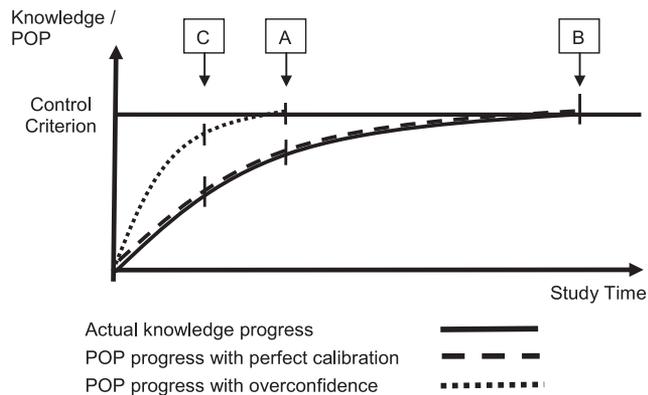
tion model. By assuming a different stopping rule that participants should adhere to, this model also implies an alternative measure of control sensitivity: the strength of the relationship between the perceived rate of knowledge gain and the decision to continue or to stop studying. Both ways of measuring control sensitivity were examined in our research.

To sum up, a metacognitive analysis of the regulation of study time yields a set of components—both cognitive and metacognitive—that potentially contribute to variance in the effectiveness of text learning, and which may differ between populations and learning conditions. In the following experiments, the MRLP methodology, which provides an integrated assessment of these components, will be used to identify and expose the possible sources of text learning differences between OSL and OPL.

Overview of Experiments

The starting point for the experiments reported in this article is the widespread preference of OPL over OSL, as discussed above. This preference was found before over a large range of age and experience levels, including young undergraduates who are used to computer use and reading texts on screen (Buzzetto-More et al., 2007).[2] On this background, we report two experiments in which we derived and compared the MLRPs of OSL and OPL. In both experiments participants studied a set of expository texts, either from a computer screen or from the printed page. Immediately after studying each text, they predicted their test performance and were tested before continuing to the next text. Because prior metacomprehension research pointed to some ambiguity regarding the type of monitoring reflected in global POP, memory of details or higher order comprehension (Maki, 1995; Pieschl, 2009; Rawson et al., 2002; Thiede, Wiley, & Griffin, in press), we asked the participants to provide two separate POPs, each targeted to one specific aspect (Kintsch, 1998).

The purpose of Experiment 1 was to examine encoding efficiency and the accuracy of metacognitive monitoring under the two study conditions, OSL and OPL. For this purpose, study time was limited to a fixed and equal amount of time per text (see Figure 1, stopping point C). By taking control of study time away from the participants, the cognitive and metacognitive components of OSL and OPL could be compared without potential contamination from the effectiveness of control decisions.

In Experiment 2, the time limit for studying each individual text was removed and the participants were free to decide how much time to allocate to each text (see Figure 1, stopping points A and B). Because metacognitive differences in the efficiency of study regulation could contribute to performance differences in Experiment 2 but not in Experiment 1, this allowed the unique contribution of self-regulation to performance differences between the two study media to be revealed and the metacognitive components underlying those differences to be examined.

Experiment 1

Perhaps the most natural account of the preference for OPL over OSL that has been examined in research so far is that display characteristics or presentation format simply make reading and writing—and hence learning—more difficult when studying text on a computer screen than when studying on paper. For example,

it might be that learning efficiency is affected by differences in reading speed or in the ease of looking back and rereading text. By this account, the primary source of media effects on learning would be perceptual-cognitive, rather than metacognitive. If so, we would expect to find a learning advantage for OPL over OSL, in terms of increased encoding efficiency, under conditions in which the allocation of study time is not under the learner's control. This issue was examined in Experiment 1. To shed additional light on potential perceptual-cognitive factors, we examined whether media differences in learning efficiency would be tied to differences in the frequency of using markup and note-taking tools (see Piolat, Olive, & Kellogg, 2005; Spencer, 2006) and whether there would be any differences in learning efficiency within the OSL group between cathode ray tube (CRT) and liquid crystal display (LCD). The display-type factor was included in the design in light of studies finding differences between the two display types that could potentially affect both objective and subjective aspects of text learning (Kong-King & Chin-Chiuan, 2000; Marmaras, Nathanael, & Zarboutis, 2008; Menozzi, Lang, Näpflin, Zeller, & Krueger, 2001; Sheedy, Subbaram, Zimmerman, & Hayes, 2005). The texts were made long enough (2–4 pages) to create potential media differences in paging-scrolling difficulty as well, though this factor was not systematically manipulated or analyzed.

Experiment 1's design and procedure allowed us to measure encoding efficiency and the accuracy of prediction of performance, in terms of both calibration bias and resolution (components 1, 2, 3, and 4 in Table 1). Those MLRP components are best measured under conditions that reduce the effects of self-regulation of study time. For this purpose, a fixed amount of study time per text was chosen (on the basis of pretesting), which allowed enough time to study the main ideas of the text, while forcing most participants to terminate their study before reaching the point at which they would naturally do so.

Method

Participants

Seventy native Hebrew-speaking undergraduate social sciences and humanities students (21 men, 49 women, mean age = 24.3 years) at the University of Haifa participated in the experiment either for payment ($15) or for course credit (11 participants). The participant recruitment notice specified that participants should not have any type of learning disability, and students who reported having learning disabilities on their personal data form were excluded from participating. The participants were randomly as-

---

[2] To reinforce the motivation for the study, we conducted a study-media preference survey (*n* = 126; 17–61 years of age). The primary target question was: "Assume that you need to read an article for serious study (such as preparing for an exam, or for a lecture you are going to give), and that the article was sent to you via the computer or that you found it on the Internet. What would you usually do? (a) print out the paper or (b) read the paper on screen." Overall, 80% of the participants reported that they would print the paper rather than read it on screen, attributing this preference to ergonomic factors such as screen glare or eyestrain, poor spatial layout, and clumsy markup and note-taking tools. Interestingly, there were no significant differences in the reported preferences of three different age groups (17–20, 21–30, and 31–61).

signed to OSL and OPL groups ($n = 35$ each). The OSL group was further divided into CRT display ($n = 17$) and LCD display ($n = 18$) conditions.

Materials

The learning materials were six expository texts dealing with various topics (e.g., the advantages of coal-based power stations compared to other energy sources, adult initiation ceremonies in various cultures, the importance of warming up before doing strenuous athletic exercise). The texts were taken from Web sites intended for reading on screen. They contained 1000–1200 words and included graphical or pictorial illustrations. When formatted and presented as Microsoft Word documents, the texts were between two and four pages long. The format and number of pages for each text were identical for on-screen and on-paper presentation. For each text, a test was devised consisting of ten four-alternative multiple-choice questions, five questions requiring memory of details, and five questions requiring higher order comprehension, with both item types intermixed. An example of a question that requires memory of details is the following: In which decade did the "coal period" start in Israel? a) 1960s; b) 1970s; c) 1980s; d) 1990s. The answer (1980s) was explicitly mentioned in the text. An example of a higher order comprehension question is as follows: The electricity production process involves a fast rotating rotor. What is the direct power source for this rotation? a) gas exhaust generated by coal combustion; b) fast flowing water; c) steam; d) hot air. The text explains that a high-temperature vapor is produced by coal combustion, which at high pressure then pushes a turbine that rotates the rotor (answer *c*).

The selection of texts and test questions for each text was based on a pretest ($n = 14$). Eight texts were used as the initial text pool. For each text, 30 questions were prepared, of which 10 questions were selected. A "good" question was defined as one that without reading the text first, the success rate was lower at least by 40% than that achieved by answering the question with the text in hand. In addition, the success rate for a text was required to be above 80% with the text in hand. This way we ensured that the questions could be answered if based on the text and that the answers were not obvious without reading the text first. If more than five good questions of the same type were found for a single text, we used the five questions that discriminated the best between answering without reading the text and answering after reading the text. The six texts with the best set of associated test questions were chosen for inclusion in the experiment.

One shorter text, of 200 words, was selected by a similar procedure and used as a practice text.

Apparatus

The computer displays were 17" CRT or LCD (MAG Technology Co. Ltd., Models 786 FD and MS776K12, respectively), both operating at 70 Hz, at a resolution of 1024 × 768. The OSL texts were presented using Microsoft Word, 2003. The font was black, 12-point, Times New Roman, at 100% scale. For the OPL condition, the same texts were printed on A4-size paper (210 mm × 297 mm; the commonly used paper size in Israel).

Procedure

The experiment was administered to groups of two to six participants at a time, all OPL or all OSL, in a room with six computer work stations. Thus, the physical room environment was the same, regardless of study media. All participants read the general instructions from a printed booklet. The only substantive difference between the sets of instructions pertained to the manner in which annotations might be made during study: OSL participants were provided with guidance about how to use the word-processing tools available in Microsoft Word, including margin comments, highlighting, underlining, and bold emphasis. All of the participants indicated that they were familiar with these basic markup tools beforehand. OPL participants were provided with a pen and a yellow highlighter as markup tools. The experimenter was present at all times.

Participants were told that they would be presented with a series of texts for study. They would be given seven minutes to read each text, during which they were allowed to make notes or mark the text for emphasis, if they wished. They were told that they would be given a multiple-choice test after each text, and that the test would include questions requiring both memory of details and higher order comprehension.

Except for the general instructions, in the OSL condition the experiment was administered entirely by computer. A master program presented the instructions at each stage for each text: opening each text in Microsoft Word for study, collecting POP judgments, presenting the multiple-choice test, and recording the answers. In the OPL condition, the same master program was used to display the instructions on the computer screen. However, instead of opening the text file for study on the computer, a window opened up on the screen, indicating the title of the text to be studied next. The participants would then take the printed text with that title from the top of the pile of texts at their station and begin reading.

At the end of the allotted study time for each text, the OSL participants saved their text file and closed Microsoft Word, whereas the OPL participants simply placed the text face down on their finished text pile. The participants then went on to the POP phase, in which separate POP judgments were elicited for memory of details and for higher order comprehension. The POP phase was administered by computer for both media conditions: POP judgments were made by dragging an arrow along a continuous scale between 25% and 100%. The question eliciting POP for memory of details was phrased as follows: "What percentage of the questions that require memory of details do you expect to answer correctly?" The same phrasing was used for the higher order comprehension POP, except that "comprehension questions" replaced the "questions that require memory of details." The instructions emphasized that the participants should evaluate their expected performance in light of the limited study time given for each text.

Immediately after the POP phase, the multiple-choice test was administered either on screen (for the OSL condition) or on paper (for the OPL conditions). Five minutes were allotted for the test, which allowed participants to answer the questions without time pressure.

The experiment began with participants reading the instruction booklet and a practice run of the entire task (study, POP, test) using the shorter practice text. The allotted study and test times for

the practice run were five minutes and three minutes, respectively. Its purpose was to familiarize the participants with the procedure and the type of test questions that would characterize the texts to follow. The set of six texts was then presented in one of two orders counterbalanced between participants. The whole procedure, including instructions and practice text, took about 90 minutes.

Results and Discussion

The main aim of this experiment was to compare the OSL and OPL conditions in terms of encoding efficiency (test performance without control of study time) and monitoring accuracy, both calibration bias and resolution. Before doing so, however, we checked whether the two potential control variables, display type (manipulated) and use of markup and note-taking tools (measured), would need to be taken into account.

Potential Control Variables

Display type. Test scores and POP judgments of the OSL group were equivalent for the two display types. Mean test score (percent correct) was 62.9 for CRT and 59.1 for LCD, $t(33) = 1.09$, $p = .28$, $d = 0.38$. Mean POP level was 69.8 for CRT and 72.2 for LCD, $t < 1$. Thus, in the following analyses both display types were combined into a single OSL condition.

Use of markup and note-taking tools. The great majority of participants (62 of 70) either marked five or six texts (20 OSL participants and 20 OPL participants) or none or one of their texts (12 OSL and 10 OPL participants). Among those who marked their texts, OSL participants used color highlighting, bold text, underlining, inserted margin comments, and added summary notes to the text; OPL participants made handwritten comments and used underlining and color-marker highlighting. The difference in the number of marked texts between OSL (3.3) and OPL (3.9) was examined by a Mann–Whitney U test, revealing no significant difference between them, $U = 547.00$, $p = .42$. Most importantly for our present concerns, when this variable was entered into the design, it did not interact with study media in any of the subsequent analyses. Therefore, it was not included in the reported analyses.

MLRP Components

Encoding efficiency. Encoding efficiency was defined earlier as the amount of knowledge gain per time unit. Based on pretesting, the preliminary knowledge level (before study) was assumed to be low and equivalent for the two groups of participants. Thus, given a fixed and equal amount of study time per text in each condition, test performance can be used as a comparable measure of encoding efficiency between the two media. As seen in Figure 2A, the average overall test score (memory of details and higher order comprehension questions combined) for the two media was virtually identical (OSL: 61.0%; OPL: 60.7%; $t < 1$), indicating equivalent encoding efficiency.

Prediction of performance. An overall POP measure was calculated as the average of the POPs for memory of details and higher order comprehension provided by each participant for each text, corresponding to the overall test scores just reported. The effect of study media on these subjective POP judgments can be seen by examining Figure 2A. Despite the equivalent level of test scores for the two study media, the combined POP was higher for OSL (71.0%) than for OPL (65.6%), $t(68) = 1.99$, $p = .05$, $d = 0.48$. Thus, although objectively there was no observed difference in encoding efficiency between the two media, the OSL participants nevertheless felt subjectively that they had learned the material better than did their OPL counterparts.

Calibration bias. To examine more directly the degree of correspondence between subjective and objective learning, calibration bias scores were calculated as the difference between the mean overall POP and test score of each participant, with a positive score indicating overconfidence and a negative score indicating underconfidence. As reflected in Figure 2A, both of the groups exhibited overconfidence. Taking into account the manner in which the POP judgments were elicited from the participants, a two-way ANOVA, Study Media × Question Type (memory of details vs. higher order comprehension) was performed on the bias scores. The main effect of study media was significant, $F(1, 68) = 2.50$, $MSE = 364.40$, $p < .05$, $\eta_p^2 = .04$, indicating greater overconfidence in the OSL condition (10.1) than in the OPL condition (5.0). A main effect of question type was also found, $F(1, 68) = 43.59$, $MSE = 87.27$, $p < .0001$, $\eta_p^2 = .39$, reflecting
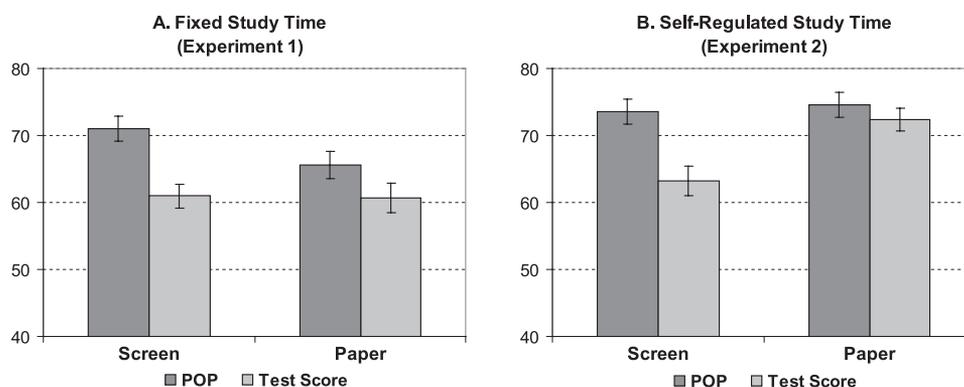

Figure 2. Mean combined prediction of performance (POP) and test scores in Experiment 1 under fixed study time (A) and in Experiment 2 under self-regulated study time (B). Error bars represent standard error of the mean.

greater overcondifence for the questions concerning higher order comprehension (test score = 60.8, POP = 73.5, calibration bias = 12.7) than for the questions concerning memory of details (test score = 60.8, POP = 63.1, calibration bias = 2.3, not significantly different from zero). There was no interaction ($F < 1$), indicating that the greater overconfidence for OSL than for OPL was not limited to a particular question type.

Resolution.   As explained earlier, whereas calibration bias reflects absolute monitoring accuracy, monitoring resolution reflects relative monitoring accuracy—the extent to which one's subjective judgments discriminate between higher and lower levels of actual performance. A common index of monitoring resolution in item-based (list-learning) memory research is the Goodman–Kruskal Gamma correlation between the metacognitive judgment and the correctness of each individual item, calculated within individuals (Nelson, 1984). This index has sometimes been extended to examine the monitoring of text comprehension, by treating each individual text as an item (e.g., Thiede et al., 2003). We did this separately for the memory of details and higher order comprehension questions and found very low correlations, with no significant difference between the two media for either memory of details (OPL = .07; OSL = .10; $t < 1$) or higher order comprehension (OPL = .11; OSL = .18; $t < 1$).

Gamma correlations become quite unstable when the number of items is small, particularly when there are "ties" on one or both variables, which further reduce the number of items that are actually included in the calculation (for other recent criticisms of gamma, see Benjamin & Diaz, 2008; Masson & Rotello, 2009). The small number of texts also precludes the use of other standard measures, such as d_a or d' (Masson & Rotello, 2009). For these reasons, we conducted an additional check by calculating the within-participant Spearman correlation between POP and actual performance on each text. Again we found very low correlations, with no significant difference between the two media for either memory of details (OPL = .08; OSL = .14; $t < 1$) or higher order comprehension (OPL = .09; OSL = .18; $t < 1$). We suspect that the low values observed for both Gamma and Spearman correlations reflect insufficient within-participant variance to meaningfully assess monitoring resolution in this experiment.[3]

To sum up: First, if the effects of technology-related factors such as display properties, mark up tools, and ease of scrolling-paging on encoding efficiency are the main source of differences between the two study media, we would expect to find a difference in test performance between OSL and OPL when study time is fixed and equated. The fact that no such difference was found counts against this possibility—a conclusion that is reinforced by the lack of effect of display type (CRT vs. LCD). Of course, these results do not rule out the possibility that display and software properties could affect learning efficiency in other contexts (cf. Kong-King & Chin-Chiuan, 2000; Menozzi et al., 2001; Sheedy et al., 2005). Second, the observed difference in calibration bias—greater overconfidence under OSL than OPL—suggests that there may be metacognitive differences between the two study media, whose effects on test performance might emerge when study time is self-regulated.

Experiment 2

As explained earlier, Experiment 2 used essentially the same materials and procedure as Experiment 1, but with one important difference: Here, the participants could decide for themselves how much time to spend on each text within a loose, global time frame. The main question was whether a difference between OSL and OPL in test performance would now emerge, attributable to differences in the effectiveness of study-time regulation between the two study media.

The combined data from the two experimental procedures (fixed study time in Experiment 1; self-regulated study time in Experiment 2) provided information regarding the MLRP components of encoding efficiency, prediction of performance, calibration bias, resolution, self-regulated performance, and self-regulated study time. In addition, to allow a more fine-grained examination of the quality of metacognitive control, information regarding ongoing changes in subjective knowledge level during study was also collected: half the participants in Experiment 2 provided "online" POP judgments during study in addition to their final POP judgments. The POP judgments elicited before and after the decision to stop studying were used to estimate the control criterion adopted by each participant and to examine control sensitivity in terms of the strength of relationship between online POP and the decision to stop studying.

Method

Participants

Seventy-four native Hebrew-speaking undergraduates without learning disabilities (mean age = 24.6 years; 24 males and 50 females) participated in the experiment, either for payment or for course credit. Half were randomly assigned to the OPL condition and half to the OSL condition. Nineteen participants in each condition received the terminal-POP procedure, whereas the remaining 18 received the online-POP procedure.

Materials and Apparatus

The same software and materials used in Experiment 1 were used again in this experiment. Because Experiment 1 yielded no effect of computer display type, only LCD displays (same as in Experiment 1) were used in this experiment.

Procedure

The procedure was similar to the one used in Experiment 1, with study media again manipulated as a between-participants variable. The participants studied each text, predicted their performance for memory of details and for higher order comprehension, and were

---

[3] Calculating gamma and Spearman correlations using the overall POP and test scores (mean of memory of details and higher order comprehension questions) yielded a similar picture: Gamma correlations averaged 0.16 for OSL and 0.05 for OPL, with no significant difference between the two study media, $t(68) = 1.04$. Spearman correlations averaged 0.19 for OSL and 0.06 for OPL, with no significant difference between the two study media, $t(68) = 1.28$, $p = .21$, $d = 0.30$.

then tested by multiple-choice questions. For the terminal-POP group, the only difference from Experiment 1 was that in this experiment participants managed their study-time allocation freely within a 90-min global time frame for studying all six texts. It was explained to the participants that this meant about 15 minutes per text, including text study, POP elicitation, and test. In the few cases (two OSL and three OPL participants) in which participants were still studying the fifth text after 70 minutes had gone by, they were asked to finish studying the text they were working on without time pressure, and the last (sixth) text was waived.

The online-POP group went through the same procedure, but in addition they were required to pause their studying every three minutes to provide a current POP judgment, for both memory of details and higher order comprehension, in addition to the terminal POP judgments provided after study was completed. The instructions emphasized that the online POPs at each point in time should take into account how much of the text had been studied so far, and how much still remained to be learned. Study interruptions for online POP were expected to prolong the overall study time, so although the same global 90-min time limit was presented in the instructions, for the online-POP group it was not enforced.

Results and Discussion

A comparison between the two methods of POP elicitation, online and terminal POP, revealed no differences in any of the dependent measures reported below. In particular, there was no interaction between POP elicitation method and study media on test scores or terminal POPs (all $Fs < 1$). Thus, until reaching the analyses of control criterion and control sensitivity, unless specified otherwise, the data analyses were collapsed across the groups.

Use of Markup and Note-Taking Tools

As in Experiment 1, most of the participants (65 of 74) either marked five or six of their texts (31 OSL and 19 OPL participants) or none or one of the texts (4 OSL and 11 OPL participants). Analysis of the number of marked texts per study media by a Mann–Whitney U test indicated that in this experiment there was a greater tendency for OSL participants (4.7) to mark their texts than for OPL participants (3.5), $U = 418.00$, $p < .01$. This finding is somewhat surprising, because people, including our survey participants, usually report that one of the reasons for their reluctance to study on screen is that the markup and note-taking tools are harder to use. Importantly, when frequency of markup was included as an additional factor, it did not interact with study media in any of the subsequent analyses.

MLRP Components

Study time. Study time was a meaningful measure only for the terminal-POP group ($n = 38$).[4] The global time limit for studying all of the texts was 90 min. The actual total study time, excluding the participants whose sixth text was waived, averaged 76.6 min, suggesting that studying was finished smoothly without any pressure. To verify that the global time-frame had not caused these participants to rush at the end of the session, a one-way ANOVA was performed to examine the effect of Serial Position (6) on study time per text. This analysis revealed a marginal effect

$F(5, 150) = 2.05$, $MSE = 2.35$, $p < .08$, $\eta_p^2 = .06$. A *post hoc* LSD test indicated that the only significant differences were between the first text, which was studied for the longest amount of time (10.1 min), and the rest of the texts (9.2 min each).

Comparison of study time per text between the two media showed that less study time was invested by OSL participants (9.1 min) than by OPL participants (10.0 min), though the difference only approached significance, $t(36) = 1.81$, $p < .08$, $d = 0.63$. This trend accords with the results for calibration bias reported below, perhaps reflecting the control consequence of overconfidence in monitoring. Note also that the participants studied each text for an average of 9.6 minutes, significantly longer than the seven minutes allowed in Experiment 1, $t(37) = 10.63$, $p < .0001$, $d = 1.72$. This reinforces our earlier assumption that the participants in Experiment 1 would generally not have reached their natural study-termination point in the fixed allotted time (see Figure 1, stopping point C).

Performance. Test scores in this experiment, under self-paced study, were lower for OSL (63.2%) than for OPL (72.3%), $t(72) = 3.34$, $p = .001$, $d = 0.79$ (see Figure 2B). To compare the pattern under self-regulated learning (Experiment 2) and fixed study time (Experiment 1), a two-way ANOVA, Experiment $\times$ Study Media, was performed on the test scores. There was a main effect of experiment, $F(1, 140) = 12.68$, $MSE = 137.66$, $p = .001$, $\eta_p^2 = .08$, a main effect of study media, $F(1, 140) = 5.12$, $MSE = 137.66$, $p < .05$, $\eta_p^2 = .04$, and a significant interaction $F(1, 140) = 5.80$, $MSE = 137.66$, $p < .05$, $\eta_p^2 = .04$. The significant interaction indicates that the advantage of OPL over OSL observed under self-regulated learning in Experiment 2 does in fact differ from the null effect under fixed study time in Experiment 1.

Prediction of performance. Figure 2B shows that despite the performance difference observed in this experiment, overall terminal POP did not differ between the two study media, $t < 1$. For the OPL participants, however, POP was higher in Experiment 2 than in Experiment 1, $t(70) = 3.28$, $p < .01$, $d = 0.78$, reflecting the increase in actual test performance in that condition. There was no difference in POP between experiments for the OSL participants, $t < 1$, corresponding to the lack of difference in test performance in that condition. This pattern suggests that POP is sensitive to differences (or lack of difference) in learning level.

Examination of online POPs provided by half of the participants ($n = 18$ in each media) allowed us to compare the subjective learning curves between OSL and OPL. Figure 3 plots the mean online POP at each elicitation point separately for each study media. The overall shape of the plots fits the theoretical learning curve presented in Figure 1. For both media, there was marked subjective progress in the initial learning stages, with decelerated progress as study continued. A two-way ANOVA, POP Elicitation

---

[4] Note that participants in the online-POP condition had to interrupt their study of each text up to four times to provide the online POP judgments. This added about two minutes to the average study time per text ($M = 11.4$ min for online-POP vs. 9.5 min for terminal-POP) as well as a substantial amount of additional variance ($SD = 1.96$ min for online-POP vs. 1.48 min for terminal-POP). Moreover, the time lost in making and switching to/from the POP judgments could not be separated from the time spent in actual study of the text. For these reasons, only the terminal-POP group was used in the analysis of self-regulated study time as a dependent measure.
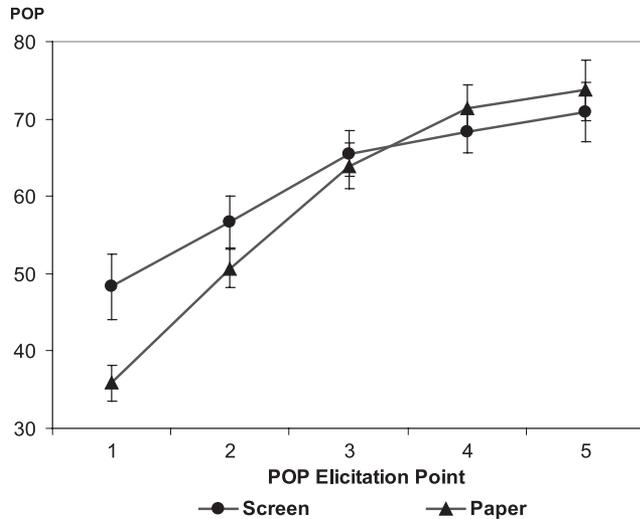
POP

Figure 3. Mean overall prediction of performance (POP) per elicitation point for OPL versus OSL in the online-POP condition of Experiment 2. Number of participants contributing to each data point was $n = 18$, except for the fifth data point for which $n = 12$ for OSL and $n = 13$ for OPL. Error bars represent standard error of the mean.

Point (4) × Study Media was performed on points 1–4 in which all of the participants had data. It revealed a main effect of elicitation point, $F(3, 102) = 116.12$, $MSE = 47.26$, $p < .0001$, $\eta_p^2 = .77$, and a significant interaction with the media, $F(3, 102) = 8.19$, $MSE = 47.26$, $p < .0001$, $\eta_p^2 = .19$. A comparison between the two study media at each elicitation point revealed a significant difference only at the first elicitation point, $t(34) = 2.54$, $p < .05$, $d = 0.87$ and nonsignificant differences at all subsequent points. OSL participants predicted their performance after three minutes to be at 48%, whereas OPL participants were more moderate in their judgments (36%). This unfounded inflation of predictions for OSL after a short fixed amount of study time accords with the POP difference observed under fixed study time in Experiment 1.

Calibration bias.   As in Experiment 1, we compared calibration bias between the two study media including question type as an additional factor. The two-way ANOVA revealed again a main effect of study media, $F(1, 72) = 6.78$, $MSE = 357.87$, $p = .01$, $\eta_p^2 = .09$, with larger calibration bias for OSL (10.4) than for OPL (2.3; not significantly different from zero). A main effect of question type was again observed, $F(1, 72) = 31.34$, $MSE = 75.69$, $p < .0001$, $\eta_p^2 = .30$, reflecting greater overconfidence for higher order comprehension (test score = 67.8; POP = 78.2; calibration bias = 10.3) than for memory of details (test score = 67.7; POP = 70.1; calibration bias = 2.3, not significantly different from zero). Finally, as in Experiment 1, there was again no interaction between the effects of study media and question type ($F < 1$), indicating that the greater overconfidence for OSL than for OPL was not limited to a particular question type.

Resolution.   As in Experiment 1, Gamma correlations were very low, yielding no difference between the study media for either question type (memory of details: OPL = .10, OSL = .21; $t < 1$; higher order comprehension: OPL = .12, OSL = .03; $t < 1$). A similar pattern was found using Spearman correlations (memory of details: OPL = .08, OSL = .21; $t(72) = 1.18$, $p = .24$, $d = 0.27$;

higher order comprehension: OPL = .07, OSL = .03; $t < 1$). Thus, as in Experiment 1, there is no evidence of media differences in monitoring resolution, though once again we suspect that the low correlations stem from low within-participant variance in POP and in actual performance between texts.

Control criterion.   The equivalent levels of terminal POP between the two study media reported earlier suggest that the same target level of knowledge may have been adopted as a control criterion. The data from the online-POP group was used to examine this possibility more stringently. According to the discrepancy reduction model, the terminal POP level for each studied text should be located just at or above the control criterion, whereas the preceding online POP, provided just before study termination, should be located below the criterion. Thus, adapting the computational procedure used by Koriat and Goldsmith (1996) to estimate the control criterion in memory reporting, we identified for each participant the POP level (average of memory of details and higher order comprehension questions) that would be below all (most) of the terminal POPs and above all (most) of the immediately preceding POPs provided for the set of texts studied by that subject. The chosen criterion estimate was the candidate POP level that maximized the "fit rate" (cf. Koriat & Goldsmith, 1996), defined as the percentage of all POPs (two times the number of studied texts) which were in fact above or below the candidate criterion level in accordance with the discrepancy reduction model. If a range of potential criteria yielded an equivalent fit rate, the midpoint of the range was used as the best point estimate.

Using this procedure, the mean estimated control criterion for participants in the OSL condition ($M = 68.5$) did not differ from that in the OPL condition ($M = 69.5$), $t < 1$. The criterion fit rates were also equivalent for the two study media (77% for OPL and 80% for OSL), $t < 1$. This finding reinforces the conclusion implied by the equivalent terminal-POP levels that the same target level of knowledge was strived for, regardless of the study media.

Control sensitivity.   The online-POP procedure also allowed control sensitivity to be compared between the two study media. According to the discrepancy reduction model, all POPs provided during the study process should be lower than the one produced after the decision to stop studying. Thus, we calculated for each participant the percentage of texts for which the highest POP (average of memory of details and higher order comprehension questions) was accompanied by the decision to stop studying. The percentage for OSL ($M = 79.4\%$) was significantly lower than for OPL ($M = 98.2\%$) $t(34) = 2.45$, $p < .05$, $d = 0.85$. By this analysis, the decision to stop studying was less consistently related to the subjective monitoring in OSL than in OPL. In fact, whereas control sensitivity was virtually at ceiling and with very little variance under OPL (16 of 18 participants yielding a sensitivity score of 100%; range: 83–100%), there was a much higher degree of interindividual variability in control sensitivity under OSL, as can be seen by the much larger standard deviation (10 of 18 participants yielding a sensitivity score of 100%; range: 0–100%).

We also analyzed control sensitivity with respect to the region of proximal learning model, which holds that change in POP, rather than the absolute level of POP, is the basis for study termination. To do so, we identified for each studied text of each participant, the minimum difference between two consecutive POPs (again using the average of memory of details and higher order comprehension questions). For each minimum difference,

the learner's decision at that point, whether to continue or to stop studying, was tabulated. The percentage of minimum differences that were accompanied by a decision to stop studying (i.e., for which the second of the two consecutive POPs was a terminal POP) was significantly lower for OSL ($M = 36.3\%$) than for OPL ($M = 59.8\%$) $t(34) = 2.77$, $p < .01$, $d = 0.95$. Thus, the application of the region of proximal learning model provides a result that converges with the result based on the discrepancy reduction model. By both analyses, the decision to stop studying was more "erratic"—less related to the monitoring output—for OSL than for OPL.

In sum, the main finding of this experiment was that in contrast to the equivalent test performance under fixed study time, performance under self-paced study was lower for OSL than for OPL. Moreover, the lower test performance of OSL was accompanied by significant overconfidence with regard to predicted performance, whereas OPL participants monitored their performance more accurately. This overconfidence difference was consistent with other differences that would be expected under the discrepancy reduction model: the somewhat shorter study time and the ensuing lower level of actual learning for OSL relative to OPL. The control criterion (norm of study) was found to be equivalent for the two study media. This suggests that the participants intended to achieve the same level of knowledge, regardless of the study media, leading us to reject a goal-setting explanation for the performance difference between OSL and OPL. Control sensitivity, on the other hand, was weaker for OSL than for OPL, implicating this factor as an additional potential source of lower OSL performance under conditions of self-regulated study.

General Discussion

The technological advances of the last few decades have led investigators to examine the potential benefits of novel methods of instruction (e.g., Chou & Liu, 2005; Chumley-Jones, Dobbie, & Alford, 2002; Macedo-Rouet, Rouet, Epstein, & Fayard, 2003; Mayer, 2003; Metcalfe, Kornell, & Son, 2007), as well as the preconditions for taking advantage of these new methodologies (Coiro, Knobel, Lankshear, & Leu, 2008; Eshet-Alkalai, 2004). However, citing a list of studies of self-regulated learning with hypermedia, Azevedo and Cromley (2004) concluded that "students have difficulties benefiting from hypermedia environments because they fail to engage in key mechanisms related to regulating their learning" (p. 523). In the present study we took a step back from the more novel aspects of the new learning technologies, examining the impact of on-screen text presentation on the more basic processes of text learning. This simplification allowed us to examine whether metacognitive learning regulation difficulties are found even in simpler computerized environments, without the extra challenges presented to the learner by advanced study techniques. We assume that the basic processes of reading and remembering expository texts on screen are essential building blocks of the more complex learning and regulatory processes that operate in more technologically sophisticated learning environments (Shapiro & Niederhauser, 2004). Thus, regardless of whatever other types of media differences in learning processes there might be, the basic differences found here should contribute to differences in almost all computer-learning environments.

In addition to shedding light on potential differences in the processes underlying text learning on screen versus on paper, an additional and independent aim of the present article was to put forward the metacognitive framework in general, and the MLRP methodology in particular, as a useful approach to the examination and analysis of objective and subjective differences in learning processes. In what follows, we first discuss how the MLRP methodology was used in the present study to uncover such differences in the underlying learning processes between the two study media, and then move on to focus on the findings themselves and their implications for the learning of texts in computerized environments.

The MLRP Methodology

The MLRP is proposed as a general methodology for analyzing study regulation in terms of its cognitive and metacognitive components, enabling the concurrent examination of the potential contributions of these components—contributions that might not be considered otherwise. The methodology is essentially a synthesis of methods based on two theoretical models of study time regulation, the discrepancy reduction model and the region of proximal learning model, and on methods developed to examine the strategic regulation of memory retrieval and reporting. All of these emphasize the causal relationships between metacognitive monitoring and control operations and the impact that these operations have on actual performance (e.g., Benjamin, Bjork, & Schwartz, 1998; Goldsmith & Koriat, 2008; Kornell & Metcalfe, 2006; Metcalfe & Finn, 2008; Nelson & Dunlosky, 1991; Thiede et al., 2003). We now discuss each MLRP component in turn (see Table 1) and consider the information that is gained by its assessment.

Encoding efficiency. In Experiment 2, under self-regulated study, test performance was lower for OSL than for OPL. However, this finding alone does not indicate the reason for the difference between the two study media. To examine potential differences in encoding efficiency, it is necessary to take away an important "degree of freedom" that learners usually have—control over the allocation of study time. This was done in Experiment 1. In that experiment, under a short and fixed study time, OSL and OPL performance was equivalent. This finding may imply equivalent learning processes in the two media, but it could also reflect the offsetting effects of differential reading speed, attention, fatigue, and many other uncontrolled factors. Whatever the underlying reasons for the equivalent encoding efficiency, the important implication is that although people are reluctant to study on screen, they can potentially do so as efficiently as on paper. This finding provides an important insight into the potential source of differences under more natural study conditions, in which learners control the amount of time allocated to each text, pointing to the role of self-regulated control of study time and the contribution of such control to learning performance.

Monitoring. In metacomprehension studies, participants are typically asked to provide POP only at the end of text learning. Under self-regulated study, however, such POPs may tap a combination of subjective encoding efficiency and study-time regulation efficiency. According to the discrepancy reduction model, learners can compensate for low assessed knowledge by investing more study time, and this is expected to bring them, at least

subjectively, to a similar level of performance (their control criterion) for all texts. Indeed, the variability of terminal POPs in Experiment 1 was larger than that of the terminal POPs in Experiment 2 (Experiment 1: Mean SD = 10.1; Experiment 2: Mean SD = 8.5, $t(142)$ = 2.05, p < .05, d = 0.35). The MLRP methodology taps the monitoring process before compensation by study-time allocation can take place, either by terminating the study early after a fixed amount of time (terminal POP; Experiment 1) or by eliciting POP early during self-regulated study (online-POP; Experiment 2). In the present study, both methods revealed a difference that was otherwise hidden: although there was no difference between study media in terminal metacognitive predictions under self-regulated study, OSL predictions were higher than OPL predictions when elicited in the early stage of study, before study-time regulation could take place.

As in the general metacomprehension literature, monitoring accuracy is examined within the MLRP methodology in both absolute (calibration bias) and relative (resolution) terms. The results pertaining to calibration bias were quite consistent between the two experiments: OSL was accompanied by a greater degree of overconfidence than OPL. Based on the discrepancy reduction model, this difference in overconfidence should have a causal effect on the allocation of study time (see Figure 1, earlier).

With regard to relative monitoring accuracy, the examination of POP resolution tends to be problematic in the context of text learning research, because of the relatively small number of judgments that can be collected and included in the calculation of the measures. In the present research, we assessed POP resolution using the Goodman-Kruskal gamma and the Spearman correlations. Both measures indicated very low levels of resolution, probably because of the small number of texts and their similar levels of difficulty.

One change from the common metacomprehension procedure, relevant to the measurement of both relative and absolute monitoring accuracy, was the elicitation of separate POP judgments for the two question types, memory of details and higher order comprehension. The separation of these POP judgments was expected to focus participants' attention on the unique aspects of each knowledge type, thereby improving monitoring accuracy. We cannot know whether the separate elicitation had any effect on POP accuracy. However, we can point to the fact that POPs for higher order comprehension were characterized by greater overconfidence than POPs for memory of details. One possible basis for such differences is that higher order comprehension judgments might reflect an evaluation of general ability (Zhao & Linderholm, 2008), whereas judgments regarding memory of details might be related more to the specific material (cf. theory-based vs. experience-based cues; Koriat, 1997). In this case, we would expect low within-participant variability in POPs for higher order comprehension relative to POPs for memory of details across the six texts studied by each participant. To examine this idea, we compared the mean within-participant standard deviation for the two question types (see Baker & Dunlosky, 2006). We found that the variability in POPs for memory of details was indeed larger than in POPs for higher order comprehension, though the difference between them was small [Experiment 1: Memory detail POP M = 11.61, SE = 0.59; higher order comprehension POP M = 10.07, SE = 0.50; $t(69)$ = 3.49, p = .001, d = 0.42. Experiment 2: Memory detail POP M = 10.12, SE = 0.57; higher order comprehension POP M = 8.33, SE = 0.60; $t(73)$ = 4.53, p < .0001, d = 0.53].

Control. Turning now to the control components, the MLRP methodology allows the level of the control criterion (norm of study under the discrepancy reduction model) to be inferred either from the terminal level of POP under the uninterrupted self-regulated study procedure, or by identifying the POP level that best differentiates the terminal level of POP from the preceding POP levels, under the online-POP procedure. No differences were found between the two study media in the level of terminal POP per se or in the criterion estimates based on the online POP procedure.

A second aspect of control that was examined is control sensitivity. As explained above, it is assumed by both of the theoretical models that study termination is tightly related to subjective monitoring, although by different stopping rules. By both of the models, control sensitivity was weaker for OSL than for OPL. As mentioned in the introduction, in the context of memory reporting, control sensitivity of healthy young adults is generally at ceiling (e.g., Koriat & Goldsmith, 1996), whereas lower control sensitivity may be diagnostic of impairment associated with schizophrenia, psychoactive drugs, and normal aging processes (see review in Goldsmith & Koriat, 2008; Pansky et al., 2009). The present finding of reduced control sensitivity under OSL in normal young adults demonstrates the potential value of this measure for exposing situational control impairments as well.

The investigation of media differences in study regulation is highly relevant for many applications. However, we also believe that the examination of these differences provides a good "case study" to highlight the general potential utility of the MLRP methodology, because it generates an intriguing situation in which equivalent groups of participants study the same set of materials but have different qualities of subjective experience. From a metacognitive perspective, this difference in subjective experience should have consequences for study regulation, which in turn should have consequences for the ultimate level of learning as measured by test performance. Thus, the comparison of the MLRPs between OSL and OPL illustrates a general approach to analyzing and examining differences in text learning processes, beyond the common focus on student characteristics and aspects of the study materials.

Metacognitive Learning Regulation on Screen Versus on Paper

After focusing on the potential contribution of the MLRP methodology to the analysis of study regulation in general, we turn now to discuss the more specific insights that can be gained by comparing the MLRPs of OSL and OPL (see Table 1, column 3). Interestingly, the common preference of OPL over OSL appears to be justified, because test performance was indeed lower for OSL under natural, self-regulated study conditions (Experiment 2). Such differences were implied in previous studies and explained in terms of display factors (e.g., Garland & Noyes, 2004) or difficulties with the use of markup and note-taking tools on screen relative to on paper (O'Hara & Sellen, 1997). However, our results discount this as the main difference between the two media: First, markup and note-taking tools were used to a similar extent in both media in Experiment 1 and even more for OSL than for OPL in

Experiment 2. Second, characteristics of the computer screen and software did not prevent participants from achieving equivalent performance levels on screen and on paper in Experiment 1 (see also Annand, 2008). The findings of no difference in encoding efficiency between OSL and OPL and the emergence of a performance difference only under self-regulated study time suggest that the efficiency of study regulation is the critical factor underlying the observed performance difference. Of course, as mentioned earlier, the generality of this conclusion will need to be examined further in future research.

Conceivably, test performance differences between OSL and OPL could also reflect differences in test media rather than in study media. However, the finding of equivalent test performance in Experiment 1 counts against the possibility that media effects on test processes are responsible for the observed performance differences in Experiment 2, whose test conditions were identical to those in Experiment 1. Nevertheless, it is worth considering the possibility that differences in test media might also affect metacognitive processes involved in the retrieval and reporting of one's answers (cf. Goldsmith & Koriat, 2008; Higham, 2007), a possibility that deserves further examination.

Turning to possible regulatory differences between the study media, one potential difference might be that participants have an initial reluctance toward studying on screen and therefore do not intend to achieve the same performance level as when studying on paper. This possibility was discounted, however, by the finding of equivalent estimated control criteria for the two media, indicating that the participants in both conditions intended to achieve similar levels of learning.

Metacognitive processes were found to differ between the two study media in two aspects. First, overconfidence was consistently greater for OSL than for OPL. A possible explanation for this difference is that the learners who studied on screen faced a more difficult learning situation. Studying difficult materials is known to increase overconfidence relative to easier materials ("hard-easy effect," Lichtenstein, Fischhoff, & Phillips, 1982). However, the hard-easy effect should reflect a pattern in which there is a large difference in performance between OSL and OPL, with a smaller difference in the subjective estimation of knowledge. In contrast, our results yielded equivalent performance, with higher POP for OSL than for OPL (in Experiment 1), indicating that OSL was not objectively harder than OPL. Thus, we conclude that the OSL overconfidence was not related to objective task difficulty.

Greater overconfidence for OSL than for OPL is especially puzzling in light of the common reluctance to study on screen (see Introduction). In fact, such reluctance might be expected to be expressed in relative underconfidence. Thus, there seems to be incongruity between the overall attitude toward OSL versus OPL and the metacognitive judgments that are made with respect to specific studied texts. This incongruity may perhaps be resolved in terms of the difference between metacognitive (first-order) and meta-metacognitive (second-order) judgments. In the context of list-learning memory tasks, for example, Dunlosky, Serra, Matvey, and Rawson (2005) asked participants to make second-order metacognitive judgments (called SOJs) which expressed their confidence in the accuracy of their first-order metacognitive judgments (JOLs). The second-order confidence judgments were found to be higher for extreme JOLs than for intermediate-level JOLs and for delayed JOLs compared with those made immediately after study.

In both cases, the second-order judgments were in fact sensitive to differences in the accuracy of the first-order (JOL) judgments. In a similar vein, it may be that one's overall subjective feeling toward OSL represents a general meta-metacognitive judgment at a more global level—in this case reflecting the perceived overall quality of one's own metacognitive monitoring and control processes when studying on screen as opposed to on paper. If learners do monitor the reliability of their own metacognitive processes and perceive these processes as generally less reliable on screen than on paper (as indicated in the present results), then this meta-metacognitive judgment could lead to a reluctance to study on screen, and would in fact appear to reflect the observed performance differences between the two media better than the first-order memory and comprehension monitoring.

The perceived unreliability of one's own monitoring during on-screen learning might also explain the second component that was found to differ between the media—control sensitivity. If one's monitoring is perceived as less reliable, one might tend less to base one's study-time control decisions on that monitoring. Following up on these ideas, the decision to print digitally presented material before study might be viewed as a meta-metacognitive control decision that transfers the study materials to the more subjectively reliable context of paper learning. This interpretation, though highly speculative, suggests the need to consider factors that affect more global (second order) self-evaluations of one's metacognitive abilities in particular contexts, which in turn may influence more specific (first order) metacognitive monitoring and control behaviors.

Why are some metacognitive processes less effective on screen? This too might perhaps be attributable in part to higher order metacognitive beliefs. Consider a related idea from the literature on age-related study deficits: It has been suggested that such deficits are related to self-referent beliefs about one's ability to effectively mobilize cognitive resources (e.g., Bandura, 1989). Older adults believe that they are less able than younger adults to recruit the needed resources when faced with a cognitive task and so may be less likely to do so (e.g., Berry, West, & Dennehey, 1989; Miller & Lachman, 1999; Stine-Morrow, Shake, Miles, & Noh, 2006). Similarly, people appear to perceive the printed-paper medium as best suited for effortful learning, whereas the electronic medium is better suited for fast and shallow reading of short texts such as news, e-mails, and forum notes (Shaikh, 2004; Spencer, 2006; Tewksbury & Althaus, 2000). The common perception of screen presentation as an information source intended for shallow messages may reduce the mobilization of cognitive resources that is needed for effective self regulation.

Research on metacomprehension has found that as people engage in more effortful processing, both performance and relative monitoring accuracy benefit (Rawson et al., 2000; Thiede et al., 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005). It may be that overcoming overconfidence bias is also facilitated by cognitive effort and deep processing, by leading to a reliance on more appropriate monitoring cues (Koriat, Lichtenstein, & Fischhoff, 1980; Sniezek, Paese, & Switzer III, 1990; Thiede, Griffin, Wiley, & Anderson, 2010).

To sum up, the results of this study point to specific metacognitive deficits in on-screen learning that do not appear to reflect difficulties in information encoding per se. In its attempt to break new ground in applying a metacognitive framework to uncover

and explain differences in learning from paper and computer screen, the present research perhaps raises more questions than it answers. Nevertheless, several potentially important implications of the findings can be specified: First, they call into question the common assumption that as long as no new technological skills are explicitly required, learners can adapt seamlessly into computerized learning environments by applying skills proven to be effective on paper (see also Garland & Noyes, 2004). Second, they raise new considerations in the development of computerized learning and testing environments, particularly when reading comprehension is involved. As just one example, when a test requires students to read a text and then answer questions, the test score is likely to reflect the effectiveness of metacognitive processes, such as time allocation and knowledge monitoring, in addition to the specific object-level knowledge or cognitive ability that is being targeted (see, e.g., Budescu & Bar-Hillel, 1993; Higham, 2007; Koriat & Goldsmith, 1998). If so, test scores may differ when the test is administered on screen versus on paper, and individual differences in the influence of test media on metacognitive effectiveness may add unwanted variance to the test scores. Third, because computerized learning environments are already ubiquitous, ways should be devised to improve the metacognitive skills of screen learners (cf. Kramarski & Dudai, 2009; Roll, Aleven, McLaren, & Koedinger, 2007). Our approach calls for a special focus of these attempts on the effectiveness of "online" metacognitive monitoring and control. Fourth, researchers who investigate study processes in general and metacomprehension in particular should pay attention to the study (or test) media as a potential intervening variable and avoid the mixing of tasks on screen and on paper as if these tasks are completely interchangeable.

Finally, in the present study we examined continuous text learning. It should be interesting to compare the effectiveness of metacognitive learning processes with hypertext and/or multimedia technologies to those of continuous text learning using the MLRP methodology. Perhaps a more active or sophisticated learning environment will enhance the effectiveness of study regulation, or on the contrary, perhaps the increased complexity and cognitive load will reduce its effectiveness. In addition, one might examine different types of learning tasks beyond continuous text learning, such as information collection and integration from multiple sources on the World Wide Web (see Britt & Gabrys, 2002; Le Bigot & Rouet, 2007; Stadtler & Bromme, 2007). More generally, this study highlights the potential utility of the metacognitive approach and MLRP methodology in identifying and revealing the source of subjective and objective differences in learning performance between different study tasks and conditions, learning materials, and learner characteristics.

References

Ackerman, R., & Goldsmith, M. (2008a). Control over grain size in memory reporting — with and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1224–1245

Ackerman, R., & Goldsmith, M. (2008b). Learning directly from screen? Oh-no, I must print it! Metacognitive analysis of digitally presented text learning. In Y. Eshet-Alkalai, A. Caspi, & N. Geri (Eds.), *Proceedings of the Chais conference on instructional technologies research 2008: 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,* 152–164.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96,* 523–535.

Baker, J., & Dunlosky, J. (2006). Does momentary accessibility influence metacomprehension judgments? The influence of study-judgment lags on accessibility effects. *Psychonomic Bulletin & Review, 13,* 60–65.

Baker, L. (1985). How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest-Pressley & G. E. Mackinnon & T. G. Waller (Eds.), *Metacognition, cognition and human performance* (pp. 155–205). New York: Academic.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology, 25,* 729–735.

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.

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73–94). New York: Psychology Press.

Berry, J. M., West, R. L., & Dennehey, D. M. (1989). Reliability and validity of the memory self-efficacy questionnaire. *Developmental Psychology, 25,* 701–713.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13,* 221–252.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings, *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Britt, M. A., & Gabrys, G. (2002). Implications of document-level literacy skills for Web site design. *Behavior Research Methods, Instruments, & Computers, 34,* 170.

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.

Budescu, D., & Bar-Hillel, M. (1993). To guess or not to guess: A decision-theoretic view of formula scoring. *Journal of Educational Measurement, 30,* 277–291.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65,* 245–281.

Buxton, B. (2008). The next Google. *Nature, 455,* 8–10.

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.

Chou, S. W., & Liu, C. H. (2005). Learning effectiveness in a Web-based virtual learning environment: A learner control perspective. *Journal of Computer Assisted Learning, 21,* 65–76.

Chumley-Jones, H. S., Dobbie, A., & Alford, C. L. (2002). Web-based learning: Sound educational method or hype? A review of the evaluation literature. *Academic Medicine, 77,* S86–S93.

Coiro, J., Knobel, M., Lankshear, C., & Leu, D. J. (Eds.), (2008). *Handbook of research on new literacies.* Mahwah, NJ: Erlbaum.

Danion, J. M., Gokalsing, E., Robert, P., Massin-Krauss, M., & Bacon, E. (2001). Defective relationship between subjective experience and behavior in schizophrenia. *American Journal of Psychiatry, 158,* 2064–2066.

Dilevko, J., & Gottlieb, L. (2002). Print sources in an electronic age: A vital part of the research process for undergraduate students. *The Journal of Academic Librarianship, 28,* 381–392.

Dillon, A. (1992). Reading from paper versus screens-a critical review of the empirical literature. *Ergonomics, 35,* 1297–1326.

Dillon, A. (1994). *Designing usable electronic text: Ergonomic aspects of human information usage.* Philadelphia: Taylor and Francis.

Dillon, A. (2002). Technologies of information: HCI and the digital library. In J. M. Carroll (Ed.), *Human-Computer Interaction in the new millennium* (pp. 457–474). Boston: Addison Wesley/ACM Press.

Dillon, A., McKnight, C., & Richardson, J. (1988). Reading from paper versus reading from screens. *The Computer Journal, 31,* 457–464.

Dillon, A., Richardson, J., & McKnight, C. (1990). The effects of display size and text splitting on reading lengthy text from screen. *Behaviour & Information Technology, 9,* 215–227.

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology, 47,* 274–296.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25,* 691–700.

Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning. *The Journal of General Psychology, 132,* 335–346.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica, 98,* 37–56.

Eshet-Alkalai, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia, 13,* 93–106.

Eshet-Alkalai, Y., & Geri, N. (2007). Does the medium affect the message? The influence of text representation format on critical thinking. *Human Systems Management, 26,* 269–279.

Garland, K. J., & Noyes, J. M. (2004). CRT monitors: Do they interfere with learning? *Behaviour & Information Technology, 23,* 43–52.

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

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition, 10,* 597–602.

Goldsmith, M., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A. Benjamin and B. Ross (Eds.), *Psychology of Learning and Motivation, Vol. 48: Memory use as skilled cognition (pp. 1–60).* San Diego, CA: Elsevier.

Hacker, D. J. (1998). Self-regulated comprehension during normal reading. In D. J. Hacker & J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in education theory and practice* (pp. 165–192). Mahwah, NJ: Erlbaum.

Higham, P. A. (2007). No Special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General, 136,* 1–22.

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

Kong-King, S., & Chin-Chiuan, L. (2000). Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics, 26,* 527–536.

Koren, D., Seidman, L. J., Poyurovsky, M., Goldsmith, M., Viksman, P., Zichel, S., & Klein, E. (2004). The neuropsychological basis of insight in first-episode schizophrenia: A pilot metacognitive study. *Schizophrenia Research, 70,* 195–202.

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., Ackerman, R., Lockl, K., & Schneider, W. (2009). The memorizing effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology, 102,* 265–279.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517.

Koriat, A., & Goldsmith, M. (1998). The role of metacognitive processes in the regulation of memory performance. In G. Mazzoni & T. O. Nelson (Eds.), *Metacognition and cognitive neuropsychology: Monitoring and control processes* (pp. 97–118). Hillsdale, NJ: Erlbaum.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Learning, Memory, 32,* 609–622.

Kramarski, B., & Dudai, V. (2009). Group-metacognitive support for online inquiry in mathematics with differential self-questioning. *Journal of Educational Computing Research, 40,* 377–404.

LaPorte, R. E., & Nath, R. (1976). Role of performance goals in prose learning. *Journal of Educational Psychology, 68,* 260–264.

Le Bigot, L., & Rouet, J. F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research, 39,* 445–470.

Le Ny, J. F., Denhiere, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarity in self-paced learning conditions. *Acta Psychologica, 36,* 280–289.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). New York: Cambridge University Press.

Lockl, K., & Schneider, W. (2004). The effects of incentives and instructions on children's allocation of study time. *European Journal of Developmental Psychology, 1,* 153–169.

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

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

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

Maki, R. H. (1998a). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition, 26,* 959–964.

Maki, R. H. (1998b). Test predictions over text material. In D. J. Hacker (Ed.), *Metacognition in educational theory and practice. The educational psychology series* (pp. 117–144). Mahwah, NJ, US: Erlbaum, Publishers.

Maki, R. H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 116–126.

Marmaras, N., Nathanael, D., & Zarboutis, N. (2008). The transition from CRT to LCD monitors: Effects on monitor placement and possible consequences in viewing distance and body postures. *International Journal of Industrial Ergonomics, 38,* 584–592.

Masson, M. E. J., & 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,* 509–527.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13,* 125–139.

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition, 18,* 196–204.

Menozzi, M., Lang, F., Näpflin, U., Zeller, C., & Krueger, H. (2001). CRT versus LCD: Effects of refresh rate, display technology and background luminance in visual performance. *Displays, 22,* 79–85.

Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based

processing heuristics? *Personality and Social Psychology Review, 2,* 100–110.

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

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132,* 530–542.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52,* 463–477.

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low risk setting. *European Journal of Cognitive Psychology, 19,* 743–768.

Miller, L. M. S., & Lachman, M. E. (1999). The sense of control and cognitive aging: Toward a model of mediational processes. *Social cognition and aging,* 17–41.

Muter, P. (1996). Interface design and optimization of reading of continuous text. In H. van Oostendorp & S. de Mul (Eds.), *Cognitive aspects of electronic text processing* (pp. 161–180). Norwood: Ablex.

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., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676–686.

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.

O'Hara, K., & Sellen, A. 1997. A comparison of reading paper and online documents. In *Proceedings of CHI'97 Human Factors in Computing Systems.* 335–342.

Pansky, A., Koriat, A., Goldsmith, M., & Pearlman-Avnion, S. (2009). Memory accuracy and distortion in old age: Cognitive, metacognitive, and neurocognitive determinants. *European Journal of Cognitive Psychology, 21,* 303–329.

Pieschl, S. (2009). Metacognitive calibration—An extended conceptualization and potential applications. *Metacognition and Learning, 4,* 3–31.

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19,* 291–312.

Pressley, M., & Ghatala, E. S. (1988). Delusions about performance on multiple-choice comprehension tests. *Reading Research Quarterly, 23,* 454–464.

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

Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition, 28,* 1004–1010.

Richardson, J., Dillon, A., & McKnight, C. (1989). The effect of window size on reading and manipulating electronic text. *Contemporary ergonomics.* London: Taylor & Francis.

Rogers, M. (2006). Ebooks struggling to find a niche. *Library Journal, 131,* 25–26.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2007). Designing for metacognition—Applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning, 2,* 125–140.

Schunk, D. H., & Zimmerman, B. J. (1994). *Self-regulation of learning and performance - Issues and educational applications.* Hillsdale, NJ: Erlbaum.

Sellen, A. J., & Harper, R. (2002). *The myth of the paperless office.* Cambridge: MIT Press.

Shaikh, D. (2004). Paper or pixels: What are people reading online? *Usability News, 6.*

Shapiro, A. M., & Niederhauser, D. (2004). Learning from hypertext: Research issues and findings. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 605–620). Mahwah, NJ: Erlbaum.

Sheedy, J. E., Subbaram, M. V., Zimmerman, A. B., & Hayes, J. R. (2005). Text legibility and the letter superiority effect. *Human factors, 47,* 797–815.

Sniezek, J. A., Paese, P. W., & Switzer III, F. S. (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46,* 264–282.

Son, L. K. (2007). Introduction: A metacognition bridge. *The European Journal of Cognitive Psychology, 19,* 481–493.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 204–221.

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.

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning, 2,* 191–210.

Stine-Morrow, E. A. L., Shake, M. C., Miles, J. R., & Noh, S. R. (2006). Adult age differences in the effects of goals on self-regulated sentence processing. *Psychology and Aging, 21,* 790.

Tewksbury, D., & Althaus, S. L. (2000). Differences in knowledge acquisition among readers of the paper and online versions of a national newspaper. *Journalism and Mass Communication Quarterly, 77,* 457–479.

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., 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,* 1267–1280.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47,* 331–362.

Thiede, K. W., Wiley, J., & Griffin, T. T. (in press). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology.*

Weaver, C. A. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 214–222.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker (Ed.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Erlbaum Inc.

Zhao, Q., & Linderholm, T. (2008). Adult metacomprehension: Judgment processes and accuracy constraints. *Educational Psychology Review, 20,* 191–206.