Text Marking: A Metacognitive Perspective

by

Zahia Marzouk

M.A., Simon Fraser University, 2014

M.A., American University Cairo, 2008

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Educational Psychology Program
Faculty of Education

SIMON FRASER UNIVERSITY

Spring 2018

Approval

Name:	**Zahia Marzouk**
Degree:	**Doctor of Philosophy**
Title:	**The Title:** Text Marking: A Metacognitive Perspective
Examining Committee:	**Chair: Dr. Lucy LeMare** Professor, Faculty of Education
	Dr. Philip H. Winne Senior Supervisor Professor, Faculty of Education
	Dr. John Nesbit Supervisor Professor, Faculty of Education
	Dr. Maureen Hoskyn Internal Examiner Professor, Faculty of Education
	Dr. Vera Woloshyn External Examiner Professor, Faculty of Education Brock University
Date Defended/Approved:	February, 23, 2018

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Abstract

Despite learners' engagement in cognitive and metacognitive processes when marking text, text marking was only examined from a cognitive perspective. Although identification of important information using textual cues and prior knowledge is a cognitive process, the decision of whether to mark or not is metacognitive. Learners use standards they create to metacognitively monitor content and decide which parts merit marking and which do not. Acknowledging the metacognitive aspect of text marking would provide a better understanding of how the study tactic works.

The current study investigates the effects of standards for metacognitively monitoring learners' interaction with text when reading and marking. The experimental design allows comparisons of performance and marking activity among groups given or not given specific criteria of content to study and mark. The research also examines standards learners use when they freely mark text as well as the probability of recall for an information segment if marked or not marked.

Learners used nstudy to mark text. nStudy is an online learning tool that allows learners to mark text and logs detailed traces of marking, and provides a description of what and how much learners' marked.

Findings show that if learners are given specific criteria to focus their learning, they do not need to mark text to process specified content. This implies that the key to efficacy of text marking is the judgment that learners engage in when deciding whether a text extract should be marked or not. Providing learners with criteria to guide marking and studying text dampens the marking of content not specified in the criteria, but it does not elevate the marking of criteria- specific content. Interestingly, learners who freely marked reported using 17 different standards when judging what to mark. Findings also show that, marking text and being engaged metacognitively in deciding what content to mark does enhance the probability of recalling marked text.

Keywords: Text marking; highlighting; underlining; metacognitive standards; reading objectives; recall; transfer

To Ahmed, my rock

Acknowledgements

I cannot express enough thanks to my supervisor Dr. Phil Winne for the continuous support of my research and study, for his patience, encouragement, and all the meaningful learning opportunities he provided. He has taught me more than I could ever give him credit for, I could not have imagined having a better advisor and mentor. I would also like to thank Dr. John Nesbit for his encouragement, empathy and very candid critical insights.

Nobody has been more important to me in my pursuit than my astounding family. To my caring, loving and supportive husband, Ahmed: my heartfelt thanks. To my daughter and son, Jude and Abdullah: thank you for the unending inspiration. To my parents, my brother, grandparents and my guardian angel Jala: thank you for believing.

I am also grateful to all my fellow labmates and friends in EdPsych lab: Mladen Rakovic, Kenny Tang, Donya Samadi, Daniel Chang, Alexandra Paztak, Jovita Vytasek, and Michael Lin for the never ending stimulating discussions and the fun we have working on projects.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Who highlights text when studying? Try asking this question in a university classroom, many hands will be raised since highlighting is university students' most preferred study tactic (Gier, Kriener & Natz-Gonzalez, 2009). An average university student reads up to 2,400 pages each semester (Caverly, Swetnam & Flippo, 1989). With this huge amount of reading university students are expected to do (Gier, Herring, Hudnell, Montoya & Kreiner, 2010), they tend to study using highlighting and underlining tactics to help them identify and isolate key concepts, as a means to focus their study (Bell & Limber, 2009). Since highlighting and underlining are conceptually the same (Fowler & Barker, 1974), for brevity, I use text marking to refer to both tactics.

Many college students report marking their textbooks increases concentration, enhances comprehension, and facilitates review (Nist & Kirby, 1986). It is easy to do, apparently requires no training, and reduces the amount of text to study and review (Blanchard & Mikkleson (1987). University study skills courses promote "*What*" and "*How*" to mark (wiki.ubc.ca/Course:COGS303). Websites advocate marking text as a method for becoming a successful learner (www.adlit.org/strategies/23332/). Text marking is so popular that a Google search of "*highlighting reading strategy*" yields 16,100,000 hits (2017 June 12 11;25). Notwithstanding its popularity, there is lack of consensus about the efficacy of text marking (Bisra, Marzouk, Guloy & Winne, 2014; Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013). Some studies showed text marking was helpful (e.g. Amer, 1994; Hayati & Shariatifar, 2009), made no differences (e.g. Johnson, 1988), and sometimes it was found to be detrimental to learning (Peterson, 1992). These contradictions prompted a meta-analysis investigating the efficacy of text marking (Bisra et al., 2014). Those findings showed learners who marked text while studying and learners who studied texts pre-marked by researchers or experts performed better on posttests than those who did not mark and those who did not study marked text ($g = 0.45$, $p \leq .001$; $g = 0.52$, $p \leq .001$).

In the research literature about study tactics, learners' marking of study materials is viewed as both an encoding and an external storage mechanism (Divesta & Gray, 1972). At encoding, it is assumed to facilitate processing information during reading text, while as an external storage technique it aids processing when learners review marked text. Text marking as an encoding mechanism is the focus of this dissertation.

When learners read and mark text they engage in cognitive and metacognitive processes. Past research, however, only use theorized cognitive perspectives to design research investigating text marking and to explain effects of text marking. The most common cognitive perspectives were levels of processing theory (Craik & Lockhart, 1972) and the von Restorff effect or isolation theory (cited in Nist & Hogrebe, 1987). The levels of processing theory focuses on text marking as an encoding mechanism. It suggests when learners actively mark information in a text, they process this information at a deeper level, which facilitates greater recall of the marked material. Anderson and colleagues' selective attention model (1982) aligns with this suggestion. It proposes when sentences are being processed, they are graded for importance. Important sentences are then noted. Marks are a visible form of such notice. Marking manifests increased attention to marked information. Marked text is therefore learnt better because of this extra processing. The von Restorff effect, on the other hand, focuses on text marking as an external storage mechanism. When reviewed, information that stands out (is marked) against a homogeneous background (unmarked text) is more likely to be remembered than other information (Nist & Hogrebe, 1987).

Deciding what to mark and what not to mark is the metacognitive component of text marking that was overlooked in earlier research of text marking. Although identification of important information, using textual cues and prior knowledge, is commonly considered a cognitive process (Abersek, 2015), the decision of whether a text deserves to be noted is metacognitive. Learners use standards they create (based on criteria they received in task instructions) to metacognitively monitor content and decide which parts merit marking and which do not.

Because text marking is the most used study tactic, universities' study skills courses offer recommendations and workshops about "what" and "how" to mark, and websites advocate marking. However, how can we recommend a study tactic when we do not know enough about how it works and especially when an integral part of it – metacognition – has been ignored in research?

1.1. The Present Study

This study posits metacognition is an integral part of text marking process. I investigate the role of standards learners use to metacognitively monitor text when reading and marking. The experimental design allows comparisons among groups given or not given specific criteria for metacognitively monitoring content to be learned and marked. A text about oil spills was created to include four categories of information: descriptions, causes, effects and remedies of oil spills. This allowed data to identify standards learners used to mark text when instructed to mark particular kinds of information. Participants in this research used software, called nStudy, while reading and marking text so traces of marking could be precisely recorded.

Specifically, this study addressed these questions:

1. Is there a difference in performance on prompted recall and transfer tasks when learners are provided criteria describing what to learn and mark text as they study vs. do not mark text?

2. Does providing criteria describing what to learn affect learners' performance on recall and transfer tasks targeting information categories specified or not specified in the given criteria?

3. Does marking affect learners' performance on recall and transfer tasks?

4. Is marking more effective when learners are provided criteria describing what to mark vs. when they freely mark using their own standards?

5. What standards do learners report using when they are not instructed to mark particular kinds of information and they mark freely?

1.2. Thesis Structure

The research questions in this study were formulated to address gaps in the literature on text marking. Chapter 2 provides an overview of theory and research on metacognition, text marking, and reading objectives research. It also examines the metacognitive aspect of text marking. Chapter 3 describes in detail the study procedures. Results are presented in Chapter 4 followed by a discussion in Chapter 5.

Chapter 2. REVIEW OF THE LITERATURE

2.1. An Overview:

2.1.1. Metacognition

Successful learners are active agents. They take control of their learning, analyze their needs, set goals and create strategies to fulfill them, all the while monitoring and evaluating their progress (Hacker, Dunlosky & Graesser, 2009; Winne, 1995). Awareness of one's cognition is a necessary prerequisite for taking charge of one's learning. This awareness and knowledge of one's cognition is part of a learner's metacognition. Metacognition is cognition about one's own cognitive processes (Baker, 2002), mental states (knowledge, feelings, and other thoughts). It is learners' "awareness of their own cognitive machinery and how the machinery works" (Meichenbaum, Burland, Gruson & Cameron, 1985, p.5). Metacognition is an important area of research in learning (Hacker et al., 2009) because findings show it accounts for approximately 17% of variance in learning among learners of different groups and backgrounds. As a comparison, intellectual abilities account for approximately 10% of variance and both share 20% of variance on various types of tasks (Veenman, Van Hout-Wolters & Afflerbach, 2006). This implies metacognition could counterbalance learners' cognitive limitations. Knowledge and beliefs about cognition, monitoring and regulating cognition are main elements of metacognition. Perhaps this explains why a metacognitive approach to learning is so popular. These elements have a bearing on any learner's work carried out in different contexts and with various learning tasks (Hacker et al., 2009).

Since the 1980s, academic learning has been regarded as a self-regulated process where learners use and direct their "thoughts, feelings and actions" to achieve learning goals (Zimmerman, 2002). The process of self-regulated learning (SRL) is comprised of sub-processes including goal setting, planning and selecting strategies, and monitoring and evaluating performance (Winne & Hadwin, 1998). The word "self" in the term SRL implies processes learners engage are "self-directed" and are influenced by "self beliefs" (Zimmerman, 2008). Monitoring and evaluating one's learning, a metacognitive process (Tobbias & Everson, 2000) is integral to self-regulated learning (Winne, 2001). According to this perspective, monitoring entails using a set of standards against which to compare the current status or condition of learning (Winne, 2001). These metacognitive standards are created by learners and, in that sense, are influenced by both external task and internal cognitive conditions (Winne & Hadwin, 1998). Task conditions include task instructions, resources available for the learner, and the learning environment. Cognitive conditions include factors such as learners' past learning experiences, personal dispositions, beliefs about knowledge and knowing, domain knowledge, and knowledge of study tactics and strategies. In much of the research literature, the role of metacognitive standards in self-regulated learning was examined over multiple study sessions. This research brings metacognitive standards to the forefront to examine their role in reading and marking text during a single study session.

2.1.2. Text Marking:

*Background**

Text marking is the most preferred study technique among college students (Gier et al, 2009). Many college students reported that marking textbooks increased concentration, enhanced comprehension, and facilitated review (Nist & Kirby, 1986). It is perceived as effortless, demands no training and minimizes material to study and review (Blanchard & Mikkleson, 1987). Apart from being favored by students, text marking is widely promoted. Study skills courses at schools and universities advocate text marking as an effective study tactic (Wade & Trathen, 1989). Study tips posted on universities' websites promote "what" and "how" to mark text to facilitate comprehension of reading texts (e.g. https://pennstatelearning.psu.edu/reading-comprehension).

Research examining the efficacy of text marking as a study tactic has encompassed multiple investigations:

1. Learner-generated marking and read-only groups:

Research findings are mixed about the benefits of text marking compared to studying without marking. Yue, Strom, Kornell and Bjork (2015) reported better performance by a marking group on recall tasks than a read-only group, especially with massed reading. Hayati and Shariatifar (2009) and Amer (1994) examined the effect of marking text on the reading comprehension of English

*This part shares ideas introduced in Bisra, K., Marzouk, Z., Guloy, S. and Winne, P. (manuscript) A Meta-Analysis of the Effects of Highlighting or Underlining while Studying.

as Foreign Language learners.

Both studies reported higher scores for marking groups compared to read-only groups. These findings were corroborated by results of Fass and Schumacher (1978), Fowler and Barker (1974), and Annis and Davis (1978) where the marking groups outperformed non-marking groups on recall tasks. These studies reported elevated performance for learner-generated marking in comparison to only reading, however, it is important to note there were some issues with the validity of experiments. For instance, Fass and Schumacher (1978) did not control for the review effect. They allowed participants to review after marking. Hayati and Shariatifar (2009) and Amer (1994) did not provide details about what the read-only groups did while treatment groups were being trained how to mark text. Moreover, when learners were requested to mark specific information, e.g. "*Mark one sentence per paragraph*" or "*Mark keywords/ phrases/ main ideas per paragraph*", their marking was not examined and no measure of treatment fidelity was reported in any of these three studies.

In contrast, Kulhavy, Dyer and Silver (1975), Idstein and Jenkins (1972), Jonassen (1984), Johnson (1988), and Reinhard, Gordon and Harris (1996) found no differences between marking and read-only groups on information retention. Marken and Marland (1979) and Hoon (1974) found no differences on reading comprehension questions.

Peterson (1992) reported no differences in information retention between marking and non-marking groups on recall; however, marking group performed worse on inferential questions when they were prompted to review only marked content. Although these studies showed no or negative effects of text marking, Johnson's (1988) in depth analysis of students' marking and recall of marked compared to non-marked text revealed elevated recall of marked information, specifically superordinate sentences, when compared to non-marked information.

2. Researcher provided marking and read-only groups:

Research comparing the performance of participants studying researcher provided marking to read-only groups also yielded contradictory results. In some studies groups studying pre-marked information outperformed groups that studied non-marked texts on recall tasks (Crouse & Idestein, 1972, Study 2; Fowler & Barker, 1974, Study1) or multiple-choice questions (Cashen & Leicht, 1970). It is important to note that Crouse and Idestein (1972) instructed participants studying pre-marked text to focus only on marked information because it would be part of the test. In doing so, they introduced another variable; test expectancy, making it difficult to attribute the elevated performance to marking alone.

These findings offer support to the Von Restorff effect which states that an item that stands out (marked) is more likely than other items to be remembered. But not all studies showed benefits of studying researcher provided marking. For instance, Coles and Foster (1975) reported no differences between groups on information retention. The same was the case in studies by Crouse and Idestein, (1972, Study 1), and Leicht and Cashen (1972). Other evidence showed studying marked text might not work for all students. Klare, Mabry and Gustafson (1955) found that studying marked text boosted performance of more able learners while it depressed performance of less able learners.

3. Learner-generated and researcher/other generated marking:

Studies comparing the effects of learner and researcher-generated marking on performance provided inconsistent findings. Some findings revealed that active engagement in marking is more beneficial than studying pre-existing marking (Fowler & Barker, 1974; Schnell & Rocchio, 1974; Rickards & August, 1975). Fowler and Barker (1974, study 1) investigated effects of (1) studying material pre-marked by a classmate, (2) actively marking and studying text, (3) studying text pre-marked by researcher, and (4) studying non-marked text. The achievement test was made up of items that targeted both marked and non-marked text. Active markers outperformed participants studying classmates' marking on items targeting marked information. And item-by-item analysis revealed that marked information was better retained than non-marked. Also, Schnell and Rocchio (1974) reported that participants generating their own marking outperformed those who studied researcher provided marking on immediate recall tasks, but they found no differences between groups on delayed tests. Rickards and August's (1975) provided more support of benefits of active participation in marking text. They examined performance on recall tasks for six groups: (1) researcher marked high-structural sentences, (2) researcher marked low-structural sentences, (3) learners requested to mark one high-structural sentence per paragraph, (4) learners requested to mark one low-structural sentence per paragraph, (5) learners free to mark one sentence of their choice per paragraph, and (6) learners studying non-marked text. No differences in marking behavior were found between the group requested to mark one sentence of their choice and the group requested to mark one high-structural sentence; both groups marked high-structural sentences. However, the group requested to mark one sentence of their choice recalled more incidental information. Except for the group marking low-structural sentences, groups generating their own marking performed better than groups studying researcher provided marking. The group marking low-structural sentences performed worse than experimenter provided low-structural marking because participants marked in an unnatural way; not as they normally do. The free marking group was observed to mark mainly high-structural sentences, the marking of low-structural group was unnatural assuming if they were marking freely they would have marked high-structural sentences like the free marking group did. Enhanced performance of learner-generated marking is credited to processing information when learners search for and select important content (Leutner, Leopold and Elzen-Rump, 2007).

In contrast to elevated performance when learners generate marking over researcher provided marking, Nist and Hogrebe (1987) found no differences in recall scores between learners studying pre-marked text and learners marking their own text when studying. However, further analysis of researcher/other-marked and learner-recalled information revealed that marked information, irrespective of whether it was important or less important, was remembered more than non-marked lending further support to the Von Restorff effect. Thompson (1980) also reported no differences in performance on open-ended recall questions between learners who marked their own text and learners who studied researcher marked texts.

Other evidence showed that learner-generated marking could be ineffective. For instance, Rickards and Denner (1979) reported that marking was ineffective when fifth graders generated their own marking. They maintained fifth graders were incapable of identifying essential information to mark as they marked more unimportant information. This may explain the ineffectiveness of learner-generated text marking. Peterson (1992) contended learners' inability to select important information to mark affects the efficiency of their marking, which could explain why researcher provided text marking could at times be more effective than learner-generated.

4. Text-marking and other study techniques:

The most common strategies compared to text marking were knowledge/concept maps and note taking. Hayati and Shariatifar (2009) and Amer (1994) compared the efficacy of text marking to knowledge maps on performance of ESL students on reading comprehension questions. The results revealed better performance for text marking groups. However, Amer (1994) reported a knowledge maps group outperformed a text marking group on summary tasks. It was hypothesized that text marking promotes success in text synthesis tasks while concept maps is more associated with success in text critique tasks (Lonka, Lindblom-YlÄnne & Maury, 1994).

Efficacy of text marking was also compared to note taking. Kobayahsi (2007) claimed the text marking tactic helps learners concentrate on reading and is not as disruptive to the process of reading as note taking. In contrast to this view, Kulhavy et al. (1975) found that note taking facilitated learners' recall more than text marking. Jonassen (1984) noted no difference between text marking and note taking group on immediate recall and recognition tasks. However on a delayed recall test, the note taking group surpassed text marking group. Hoon (1974) also showed no differences in learners' information retention between text marking and note taking groups, as did Ayer and Milson (1993) with seventh graders and Wilder (1982) on immediate and delayed recall and recognition tasks. Issues with treatment fidelity were observed in Hoon (1974) and Wilder (1982). Hoon (1974) examined the effect of different study tactics (marking, note taking and read-only) in a within-subject design but did not describe the order of treatments. Treatment diffusion is a potential confound. Wilder (1982) requested the marking group to first read text, reread and then mark where one of the comparison groups was rereading. This is a confounded effect.

Text marking was also compared to outlining and summarizing in Stordahl and Christensen's study (1956). No differences between groups were found on immediate and delayed reading comprehension tests.

5. Effect of training on text marking:

Following undetected differences between marking and non-marking groups on recall tasks, Peterson (1992) maintained that the efficacy of text marking is dependent on students' ability to identify important information to be marked. Nist and Simpson (1988) remarked that in most studies examining the efficiency of text marking, students were not given any training on how and what to mark. Based on the literature reviewed, training learners on how to use the study tactic led to better identification of important information which consequently affected performance. Glover, Zimmer and Filbeck (1980) carried out training to enable students to identify the semantic base of reading texts over a period of 7 weeks. Participants practiced marking text then were provided immediate feedback about their marking. They reported enhanced performance on reading comprehension questions, as well as an increase in correct marking of text and a decrease in marking of extraneous information after training sessions. Research that trained participants on how and what to mark, irrespective of how short the training was, reported elevated performance for marking groups. For example, Hayati and Shariatifar (2009) provided 60 minutes of training to participants before they studied and marked text. The training included suggestions on when to mark, how and what. Results of Amer (1994) are also in favor of training. In this study participants were trained in 90 minute sessions once a week for five weeks. Participants were given four steps adapted from Smith (1985) to follow when attempting to mark text. Crewe (1969) also trained participants on how to mark but details of training content and procedure were not provided.

Research comparing trained and non-trained marking groups revealed mixed results. For instance, following a 45 minute training session that included tips with examples based on Pauk (1974) to follow when marking text, Craik and Martin (1980) reported no detectable effects on recall between a training group and a no training group. The training group did recall most of the tips on marking provided in training. Craik and Martin (1980) maintain that providing students with helpful tips does not mean that they will apply these tips when studying. It is crucial to point out that Craik and Martin (1980) trained participants to mark text and write notes in the margins. This tactic is not text marking but rather a mixed method.

Schnell and Rochioo (1974) found that a marking without training group performed better than a group trained to mark text on immediate recall. However, the trained marking group surpassed the marking without training group on delayed tasks. Moreover, participants in the marking without training group performed better than a group studying text with researcher provided marking on immediate recall tasks, but no differences were detected on delayed tasks.

Leutner et al (2007) examined the effects of two different training treatments, text marking learning strategy versus text marking learning strategy and self-regulation. They found that learners trained in the learning strategy outperformed those with no training and participants in the learning strategy and self-regulation group performed far better than those trained in the learning strategy only.

6. Individual differences and text marking:

Marking behavior of different ability students was investigated. Wade and Trathan (1989) reported no differences in marking behavior for low and high ability university students. Both high and low ability students recognized and recalled important information irrespective of whether it was marked or not. Contrary to this finding, Bell and Limber (2009) surveyed marked textbooks of university high and low skill students and found highly skilled students marked less text than low skilled. They also reported that low skill readers were not as capable of identifying crucial information. This inability to identify important information increased the probability of getting lower grades.

Success of text marking by readers with different skill levels was also researched. For instance, Blanchard and Mikkelson (1987) reported that better readers benefited more from marking text than less able readers. Stordahl and Christen (1956) revealed similar findings. Perhaps less able readers are not capable of finding important information, which consequently limits their performance as proposed by Bell and Limber (2009). There is also evidence that even when learners are provided marked text; more able learners profit more than less able ones (Klare et al, 1955; Crouse & Idestein, 1972).

In a study by Annis and Davis (1978), prior knowledge and topic familiarity were also examined with regard to text marking. They reported text marking was more effective when the topic was familiar. Marzouk, Winne and Nesbit (manuscript) also found prior knowledge mediated the effect of marking on both recall and transfer tasks. Learners with more prior knowledge benefitted more from marking text.

The relationship between text marking and learners' motivation also has been examined. In a study by Fass and Schumacher (1978), motivated learners marked more effectively where marked information included answers to questions on recognition test than unmotivated ones.

7. Inappropriate text marking:

The increased consumption of used textbooks, which often have markings, prompted researchers to investigate the influence of reading text that is inappropriately marked, i.e., marked text is irrelevant to achievement test questions. Silvers and Kriener (1997, study 1) examined effects of pre-existing appropriate (relevant to test questions) and inappropriate marking (irrelevant to test questions) on students' reading comprehension. The inappropriate marking group performed worse than the appropriate marking and no-marking groups. No differences were detected between appropriate marking and no-marking groups. Even when participants were warned about inappropriate pre-existing marking, they could not ignore it and consequently their performance suffered (Silvers & Kriener, 1997, study 2). Grier et al (2009) found similar results. They examined effects of inappropriate marking on learners' reading comprehension and metacognitive processing represented by judgment of learning. Performance on reading comprehension suffered when students studied inappropriately marked texts of low or high difficulty. No differences were detected between no-marking and appropriate marking groups. Inappropriate marking also hindered learners' ability to metacognitively monitor their learning. But research by Geir et al (2010) found that students could overcome effects of inappropriate marking by making their own marking using different colors.

Factors possibly affecting efficacy of text-marking

A text marking meta-analysis (manuscript) identified several factors that could affect the efficacy of text marking.

Immediate and Delayed Performance

Research examined how text marking affected learners' immediate performance compared to delayed performance. Stordahl and Christensen's (1956) found no differences on immediate and delayed reading comprehension tasks in a learner-generated marking group, as did Wilder (1982) on immediate and delayed recall and recognition tasks, and Hartley, Bartlett and Branthwaite (1980) for researcher provided marking. Hartley et al. (1980) reported that high and average ability sixth graders studying marked text outperformed participants' who read non-marked passages on both immediate and delayed recall.

However, Thompson (1980) reported better performance for marking groups on immediate recall compared to delayed recall. In contrast, Jonassen (1984), found better retention of information after a one-week delay compared to immediate recall. Schnell and Rocchio (1974) reported that participants generating their own marking outperformed those who studied researcher provided marking on immediate recall tasks, but they found no differences between groups on delayed tests.

Other studies investigated effects of delay preceded by review. Crewe (1969) found elevated effects of delayed recall performance when preceded by a review after four weeks delay. Fowler and Barker (1974) also showed that learners scored higher on questions related to marked parts after one-week delay. Marked text was better retained than unmarked text when Nist and Hogrebe (1987) measured recall after a one-day delay when participants were allowed 10 min for review. However, Peterson (1992) reported a negative effect on inferential questions when there was a one week-delay followed by review.

Review and no review

Text marking is viewed not only as an encoding mechanism but also as an external storage when learners mark text with the intention of reviewing it at a later time (Divesta & Gray (1972). Kornell and Bjork (2007) reported 60% of 472 undergraduates use marked text as a guide for future studying.

Some studies have investigated the efficiency of text marking as a review strategy. For instance, Idestein and Jenkins (1972) examined effect of marking on students' retention of information after a one-week delay with nine or four and a half minutes given for review. No effect of marking was detected but there was a review effect. Both marking and non-marking groups given nine minutes to review performed better on completion questions than learners allowed four and a half minutes. Yue et al (2015) examined the effects of marking followed by either a spaced and massed review. Results showed that massed review boosted the benefits of marking text.

One of the main reasons for the popularity of text marking is it reduces the amount of text to study and review (Blanchard & Mikkleson (1987). But is this reduction useful to learning? Results reported by Idestein and Jenkins (1972) and Hoon (1974) revealed no detectable effect of reviewing marked text on learners' performance. Johnson (1988) also found no effect of review when they measured recall after review of text with markings. Peterson (1992) even argued that over-reliance on reviewing only marked text might impede deep processing of the whole text and prevent the construction of internal connections among ideas.

Text importance, marking and recall

Analysis of recall of marked compared to unmarked text showed that marked text had a better chance of being recalled than non-marked (Winne et al., 2017; Winne, Marzouk, Ram, Nesbit and Truscott, 2015; Idstein, 1975) irrespective of learners' reading ability or study time (Blanchard & Mikkleson, 1987) or whether marked information was important or less important (Nist & Hogrebe, 1987). Moreover, Cashen and Leicht (1970) found that both marked and text-adjacent-to-marked information were better retained than non-marked information.

Prompted by research that suggests higher-level information in a reading text serves an assimilative function for subsuming lower-level information (Ausubel, 1968), some research examined the effects of marking high-level sentences on information recall. High-level sentences are sentences that are important to the overall meaning of a text; these sentences are abstract and high in "semantic or structural importance." In contrast, low-level sentences are more specific and provide examples and elaborate higher-level statements (Rickards & August, 1975, p. 860).

Rickards and August (1975) reported group instructed to freely mark one sentence per paragraph, marked high-level sentences and recalled twice as much unmarked information than group studying high level sentences generated by researchers. However, Johnson (1988) found that learners in a text marking group recalled more superordinate sentences, whether marked or not, than subordinate sentences. Wade and Trathan (1989) also showed that important information is learnt regardless of being marked or not.

Measures of achievement

Most studies examining the benefits of text marking measured recall or reading comprehension. Research measuring recall reported mixed results. For instance, Crouse and Idestein (1972) reported researcher generated marked text was better recalled than non-marked, as did Nist and Hogrebe (1987), Fowler and Barker (1974), Coles and Foster's (1975, study1) and Rickards and Denner's (1979). Some studies examining learner-generated marking also reported elevated recall for marking groups compared to read-only groups (Fowler & Barker, 1974). Others found no differences (Peterson, 1992; Johnson, 1988; Thompson, 1980). Peterson (1992) suggested that learners' inability to identify important information affects the success of their marking.

Research measuring reading comprehension also had inconsistent results. Grier et al (2009) and Silvers and Kreiner (1997) found no differences in comprehension scores between a group studying researcher-marked text and a read-only group. Hayati and Shariatifar (2009) reported better performance for learner generated marking than read-only, as did Amer (1994). However, other studies found no differences (Stordhl & Christensen, 1956; Hoon, 1974). Peterson (1992) showed lower performance for marking groups on inferential questions after a one-week delay with review of only marked segments. It seems that reviewing only marked information, removed from the context of the reading material, has a negative effect on making inferences.

There is little text-marking research measuring transfer. One study that explored a relation between transfer and text marking is Marzouk et al. (manuscript). We found marking information that is incidental to transfer posttest items predicted better transfer. It may be that material incidental to information central to transfer helps learners elaborate central information in some way that benefits conceptual understanding which, in turn, helps to answer a transfer posttest item.

Instructions

Research examining the benefits of learners' generated text marking either instructed text marking groups to mark specific text or allowed them to freely mark. Studies requesting learners to mark freely revealed either no effect or negative effects of marking on learners' performance compared to reading-only or note taking groups. For instance, no differences were reported between marking and read-only groups on reading comprehension questions (Stordahl & Christensen, 1956) and on recall (Peterson, 1992). Other studies showed a negative effect on inferential questions tasks when a marking group was compared to a read-only group (Peterson, 1992), and on recall when a marking group was compared to researcher provided marking (Nist & Hogrebe, 1987).

Studies providing specific instructions on what to mark reported contradictory results. For example, research instructing learners to "*Mark one sentence per paragraph*" (Johnson, 1988) or "*Mark three lines on each page*" (Kulhavy et al, 1975) revealed no differences between marking and read-only on recall, no differences for open-ended recall between learner-generated marking and researcher provided marking (Thompson, 1980), or showed learner-generated marking was not effective on recall tasks (Rickards & Denner, 1979).

Fass and Schumacher's (1978), on the other hand, required a marking group to "*Mark key words or phrases*." The group engaging in learner-generated marking outperformed a read-only group on recall.

Summary

Although research reported contradictory findings regarding efficacy of text marking, a meta- analysis by Bisra et al (2014) revealed that actively marking or studying marked text enhance learners' performance on recall and reading comprehension tasks.

Several observations could be made about the design of research examining text marking and analysis of learners' marked text:

1. Most research ignored the role of reading objectives on learners' marking and processing of text.
2. Learners' marking was rarely examined. Very few studies analyzed and reported on information marked versus non-marked, and compared recall of marked to non-marked text.

3. Text marking research mainly measured recall and reading comprehension, very little measured transfer.

4. With regard to marking and recall of low and high level sentences, some findings showed that marked sentences were recalled regardless of their status. Other findings revealed that high level sentences were recalled whether marked or not.

5. The metacognitive component of the text marking study tactic was not examined in research. There are only two studies that mentioned metacognition in context of text marking: Leutner et al (2007) and Yue et al (2015). Leutner et al (2007) requested learners to reflect on their marking after they read and marked text. Although this research examined the influence of reflection (a metacognitive component) as part of both text marking and SRL strategy, it did not examine the metacognitive aspect of text marking while reading and marking text. Yue et al (2015) simply surveyed learners' metacognitive beliefs about text marking study tactic.

The current study attends to gaps in earlier text marking research by exploring the influence of instructions on learners' marking behavior and their performance measured by recall and transfer posttests. More details are provided in the following sections.

Theoretical perspectives

Although text marking involves both cognitive and metacognitive processes, it has been mainly researched from a cognitive perspective. Adopting Winne 's (2001) "*if-then*" view of a study tactic sheds light on what takes place when a reader is interacting with information. According to this view, if a learner judges a set of criteria is satisfied then a study tactic is applied. The judgment component of this sequence is a metacognitive act because it involves learner's thinking about and using self-created standards to guide learner's cognition about the text (Winne, 2001). Applying this definition to text marking, when learners are reading and marking text they use metacognitive criteria to identify which information merits marking. The metacognitive component of text marking was not examined by earlier research. Later, more details regarding this point will be provided. Immediately below, I briefly review the cognitive account offered in prior research.

In studies reporting better retention of marked information, the effectiveness of text marking was explained mainly using two theoretical perspectives: levels of processing theory (Craik & Lockhart, 1972) and the Von Restorff effect (cited in Nist & Hogrebe, 1987), **The Levels of Processing Theory** focuses on operations at input where information is being processed when reading and marking text, i.e., marking as an encoding mechanism. According to this perspective, encoding of information undergoes different levels of analysis that begins with sensory analysis of a stimulus, a low level processing, followed by matching features of stimulus to patterns in existing knowledge and finally deep semantic analysis where Craik (2002) refers to as "analysis of meaning, inference and implication" (p. 309). Deep semantic analysis is a higher level processing that is associated with higher levels of retention. Findings of several studies supported the levels of processing theory (e.g. Johnson, 1988; Amer, 1994). **The Von Restorff or isolation theory**, on the other hand, focuses on processing information when the marked text is reviewed−text marking as an external storage mechanism (cited in Nist & Hogrebe, 1987). It theorizes that an item that stands out, such as by marking is more likely to be remembered than information in the background (Nist & Hogrebe, 1987). Findings of studies examining learners' studying of pre-marked text support this theory (Cashen & Leicht's, 1972; Leicht & Cashen's, 1970). The Von Restorff effect is beyond the scope of this research since the current study examines text marking as an encoding mechanism.

According to the Levels of Processing Theory, as commonly used in the literature, when learners search for, select and are actively engaged with information they read, they are deeply processing this information (Nist & Hogrebe, 1987). Deep processing, in turn, facilitates storage and recall of information (Leutner et al, 2007). I posit this interpretation is cursory for two reasons. First, it does not explain how search and selection are metacognitively guided. Is all information read and processed, then judged whether to be marked? No specific operations are identified to explain how this happens. Second, although it is safe to claim well-recalled information was deeply processed (Baddeley, 1978), the assumption that marked text is processed more deeply than non-marked text falls short of explaining why unmarked text, i.e., incidental learning, is also remembered (e.g., Fowler & Barker, 1974; Rickards & August, 1975). Acknowledging and investigating the metacognitive component of text marking may shed more light on these issues.

2.1.3. Reading Objectives

Having objectives is important for learning as they define "where you are headed and how to demonstrate when you have arrived" (Kaufman, 2000, p. 44). In educational settings, learners study for a purpose (McCrudden & Schraw, 2007), This purpose affects how they process text (Elliot, 1999) and consequently their learning. Reading objectives are either set by the learner e.g. *I need to study causes of oil spills*, or by the task instruction, e.g. *Study causes and remedies of oil spills* (Tilstra & McMaster, 2013). Either way, the objective influences a learner's choice of information that merits attention which in turn impacts the mental representation of the text that the learner forms (Tilstra & McMaster, 2013).

Bråten and Samuelstuen (2004) stated that when learners are given clear instructions of what they are expected to know while reading, before reading a text, they are theorized to use these objectives to guide processing of the text, specifically to identify idea units they need to learn. This in turn influences their comprehension of the text. In a meta-analysis examining intentional and incidental learning of reading texts, Klauer (1984) reported instructional objectives boost learning information relevant to the stated objectives (i.e. intentional learning) but undermine learning of irrelevant information (i.e. incidental learning). Klauer theorized instructional objectives provided in advance of studying govern intentions; intentions, in turn, guide learners' choice of what deserves attention when studying text.

Studies indicated task instructions influence operations learners engage during learning. Blischak and Challis (1994) pointed out students usually ask about the type of test questions (i.e. recall or recognition) they should expect when preparing for an exam. This suggests students intend to modify their study strategies based on the expected test format. Entwistle and Entwistle (2003) interviewed university students regarding their study strategies in preparation for final exams. They found students report adjusting their study strategies according to teachers' expectations of their performance in final exam. However, few studies examined how test expectations affected learners' study strategies (Abdel Fatah, 2011). For instance, Feldt and Ray (1989) investigated how test expectancy affected learners' choice of study strategy and how that choice influenced performance. They reported most learners' expecting multiple-choice questions underlined and reread text, and did not take notes. Out of 37 learners, 17 studied similarly to learners expecting free recall, while nine learners expecting free recall reread, underlined and took notes. No detectable differences were observed on test scores between groups. Abdel Fatah (2011) examined whether undergraduates adapt their study strategies to align with a task's cognitive demands and how this adaptation influenced performance. One group was told to expect surface level questions while a second group was instructed to expect deep level questions. Learners expecting deep level questions reported they stressed deep level study strategies more than learners who expected surface level questions, while learners expecting surface level questions reported using surface level strategies. Abdel Fatah also reported a mediating effect of study strategies intervening between learners' expectations of type of questions and performance on these questions. Moreover, Surber (1992) examined the effect of text length, test type expectation and subject matter on how learners study and highlight text. Learners studying shorter texts (around 700 words) highlighted more than learners studying longer texts (around 4,600 words).

Providing reading objectives through task instructions for learners affects not only their choice of study tactics (Brâten & Samuelstuen, 2004) but also the way they use a study tactic. Marzouk et al (manuscript) found providing different reading prompts before learners studied and marked text affected text marking behavior. Groups oriented to prepare for a recall task marked less text and fewer idea units representing examples, main ideas and reasons than groups oriented to expect a transfer task. Instructions not only affected the quantity of marking but also categories of information learners chose to mark, which consequently affected performance on transfer tasks.

The significance of task instructions is highlighted in Winne and Hadwin's (1998) model of self-regulated learning (SRL). The model presents four phases of studying: (a) task definition, where the learner forms a definition of her understanding of what the learning task entails; (b) goal setting and planning, where the learner sets specific goals to accomplish the task and plans how to achieve these goals using set of strategies; (c) enacting studying tactics, where the learner uses the strategies she identified in earlier phase and (d) adaptations, where the learner decides to make adaptations to her belief system, motivation and strategies. According to the model, learners adapt and regulate cognitive strategies to align with task demands.

Task instructions that include reading objectives are perceived to "explicitly describe what the learners should know or be able to do at the completion of instruction" (Smith & Regan, 1993, p.65). Since objectives signal to learners "why, how and what to read" (McCruden & Sparks, 2009), learners, each with unique cognitive conditions, use reading objectives to construct personal standards for judging which information is worthy of attention. While reading and studying, learners use these metacognitive standards to guide search and selection processes to locate information that merits attention (Winne & Hadwin, 1998).

Although research findings support the influence of task instructions on learners' choice and use of study strategies, most text marking research ignored this. To clarify, in some studies where marking groups outperformed a non-marking group, the marking groups were given an objective for reading and marking while the non-marking groups were not. For instance, in Hayati and Shariatifar's (2009) and Amer's (1994) studies, marking groups were requested to "*read and mark important information"* while non-marking groups were only asked to read text. However, studies that reported no statistically detectable differences between marking and non-marking groups prompted non-marking groups to make some sort of judgment while reading and studying. A case in point is Jonassen (1984). The marking group was asked to "*underline sentences that were important to the overall meaning of the passage*" while the non-marking group was told to "*read the passage and reread any section not comprehended*" (Jonassen, 1984). Kulhavy et al (1975) requested marking groups to "*underline up to three lines*" while non-marking were asked to "*study the material carefully*". In these examples, differences in instructions may affect standards used in metacognitive monitoring what to mark, which, in turn, accounts for findings. Thus, while acknowledging the cognitive component of text marking, my study examines the role of standards for metacognitively monitoring learners' processing of text while reading and marking.

2.2. Metacognitive Aspect of Text Marking

Reading is a complex cognitive process involving various sub processes (Mckeown & Beck, 2009) some of which are metacognitive. Choosing when to use a reading strategy or tactic, how to use it, evaluating and monitoring comprehension, and controlling strategy and use are all examples of metacognitive processes taking place while reading (Baker & Brown, 1984). Decisions about which content to study is also metacognitive (Kornel & Bjork, 2007).

To map cognitive and metacognitive processes occurring when learners read and mark text, I put forward a model of what happens when learners mark text. My model is guided by Craik and Lockhart's (1972) encoding theory of levels of processing. It is important to note, however, that the focus of this research is beyond testing this model. The model is provided at this point to help situate the research questions.

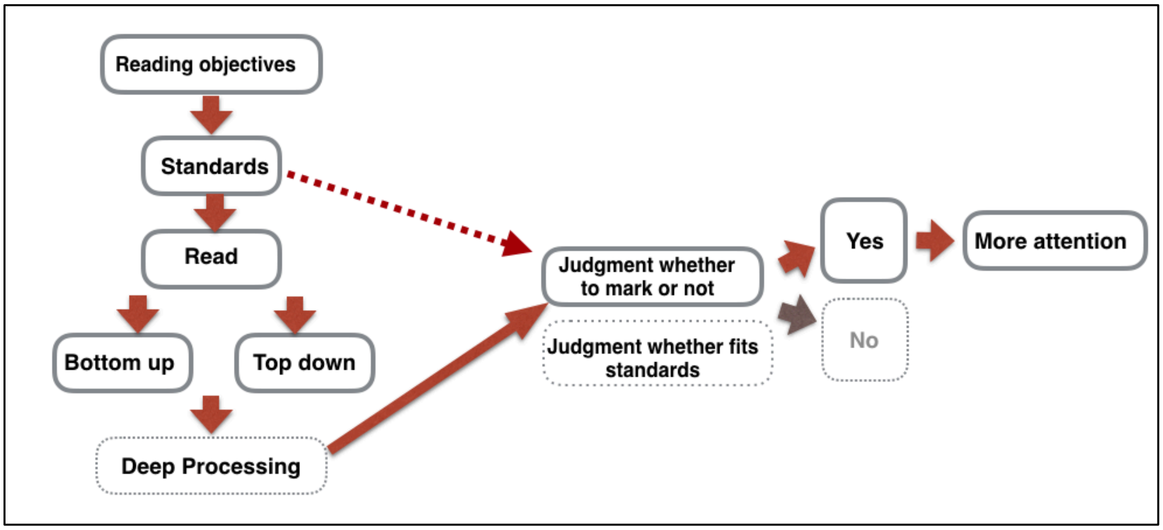


Figure 2.1. Text marking model. Cognitive and metacognitive processes during reading and marking text

According to the proposed model, learners first are given an objective (e.g., "*Read the following text and mark important information*." OR "*Read and mark effects of oil spills"*. As discussed earlier, learners use these objectives to create standards to guide judgment when reading text about which text to mark.

Then learners read the text interactively using stimulus-driven bottom-up processing and conceptually driven top down processing grounded in prior knowledge (Craik & Lockhart, 1972). The model suggests that, up to this point, all the text is processed in approximately similar ways. This may explain recall of some unmarked text reported in studies examining text marking (e.g. Fowler & Barker, 1974).

During study, learners use metacognitive standards created in relation to objectives to *judge whether to mark*. If learners are not required to mark overtly, learners still use these standards to metacognitively judge whether a text segment is worthy of attention to improve recall.

Information *worthy of attention* does not necessarily mean it is *important*. Since, identifying important information in text is widely regarded as a cognitive process (Abersek, 2015), a distinction between two key concepts: *importance* and *relevance*, is needed at this point. As put forward by McCrudden and Schraw (2007) "relevance is the degree to which a text segment is germane to a specific task or goal, whereas importance is the degree to which a segment contains essential information needed to understand a text" (p.114). *Important* segments in text often are cued by the author (e.g., by typographical cues, order of presentation). Thus importance is text-related. *Relevance*, on the other hand, is determined by learner's objectives or standards. It is a text-external phenomenon (McCrudden & Schraw, 2007). A relevant text segment does not need to be important. Schraw, Wade and Kardash (1993) demonstrated the difference between importance and relevance. A group of participants rated each text segment in a narrative for importance using three ratings: low, medium and high importance. Then two other groups were requested to read the same narrative, one group from the perspective of a burglar and the other group as a homeowner. Each group was instructed to distinguish low, medium and high relevance text segments. Participants with the burglar perspective recalled segments related to the burglary even though these segments were of low importance. However, text segments rated low relevance but high importance were recalled better than low relevance and low importance. Schraw et al. (1993) inferred when readers are monitoring text to remember, they use a text based importance by default but, when they generate criteria for identifying more and less relevant information, they change to a relevance criterion to evaluate text segments and guide their processing.

Adopting McCrudden and Schraw's (2007) concept of relevance, learners generate metacognitive standards in relation to cognitive and task conditions, then use those standards to judge whether a text segment deserves marking. Marked text signals a learner judged it *relevant* (text-external). In other words, the process of judging whether a text segment merits attention has both cognitive and metacognitive components. The cognitive component includes text encoding processes and accessing prior knowledge to comprehend what is read; the metacognitive part involves (a) monitoring, applying and continuously adjusting standards, and (b) controlling processes leading to mark text or not. Following Nelson's (1996) meta and object levels, the meta level monitors information from the object level (outcome of integration between encoded text and prior knowledge) then exerts control by informing the object level what needs to be done, in this case, whether a text should be marked or not.

Returning to the proposed model, earlier researchers interpreted when learners mark text, there are two possible reasons why marked text was better remembered than non-marked text: First, the decision to mark (specifically select) text may classify and differentiate it in a meaningful way (Goldstein & Chance, 1971). This decision to mark is the *metacognitive* component of text marking. Second, to mark text, as learners use a marker or cursor to select text, they may reread it. This may rehearse the marked text. Rehearsal, a possible *cognitive* component of text marking, has been found to increase learning in research unrelated to text marking (Craik & Lockhart, 1972). Providing identical criteria to two groups, one requested to mark and the other do not mark, allows examining the metacognitive and cognitive components of text marking.

Testing transfer was rarely observed in previous text marking research and findings where recall was measured were inconsistent. Thus, I was curious to examine the association between learners' text marking and performance on both recall and transfer items In the context of the research literature, I believe theory could be advanced if the following research questions were investigated:

1. Is there a difference in performance on prompted recall and transfer tasks when learners are provided criteria describing what to learn and mark text as they study vs. do not mark text?

2. Does providing criteria describing what to learn affect learners' performance on recall and transfer tasks targeting information categories specified or not specified in the given criteria?

3. Does marking affect learners' performance on recall and transfer tasks?

4. Is marking more effective when learners are provided criteria describing what to mark vs. when they freely mark using their own standards?

5. What standards do learners report using when they are not instructed to mark particular kinds of information and they mark freely?

Chapter 3. METHOD

3.1. Participants

Participants were 66 undergraduate students, 42 females and 24 males with various disciplinary majors attending a university in Western Canada. Ages ranged from 18 to 38 years (M =21, SD=3.43). All participants were recruited via an advertisement posted in a busy spot on campus and were compensated $12 for taking part in the study.

3.2. Materials

A text was created consisting of 90 sentences (1802 words) about oil spills. The text had three main sections: an introduction describing oil spills in general; descriptions, causes, effects and remedies of marine oil spills; and descriptions, causes, effects and suggested remedies of land based oil spills.

Each sentence in the text was categorized as belonging to one of four categories except for 3 sentences that were segmented and segments assigned a category each. (a) Description was information describing what is oil spill, what is not as well as information about the current status of oil spills (16 sentences). (b) Cause was information about causes and sources of both marine and land based oil spills (21 sentences); (c) Effect was information about effects of oil spills on the environment (25 sentences). (d) Remedy was information about ways to deal with oil spills (24 sentences). Seven sentences were elaborations; added details. Each was given the same label as the category it elaborated; two causes-elaborations, three effects-elaborations, and two remedies-elaboration. A sample sentence that was divided into two categories was: *"Improper waste management of oil facilities located on agricultural lands (cause) could lead to soil and water contamination (effect)."*

In reading materials about a problem, content is often blocked; sentences describing the problem come at the beginning of a block followed by sentences expressing causes, then sentences describing effects and finally sentences introducing remedies. Because such a block presentation may signal category type the reading text, used here dispersed sentences types throughout the text. Figure 3.1 illustrates the pattern used in this study.

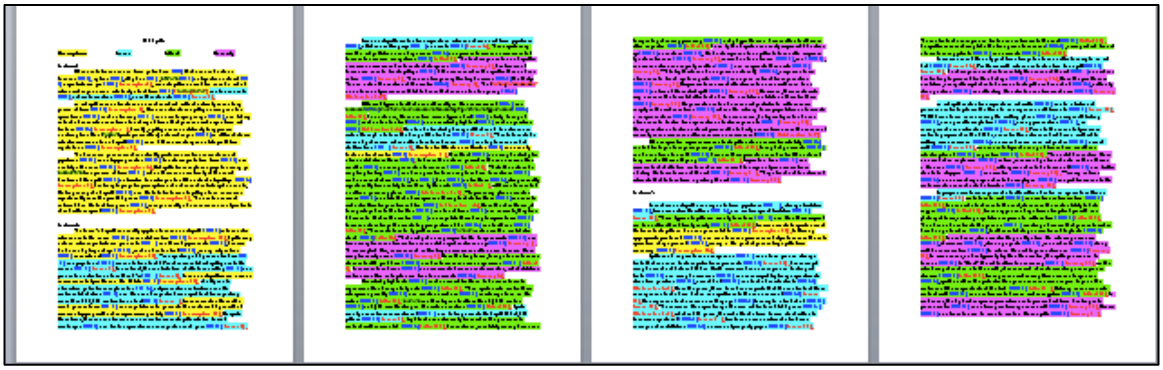


Figure 3.1 Color-coded screenshot that shows how the four categories are dispersed in the reading text

To lessen the likelihood participants relied on superficial cues when deciding what to mark, words and phrases that can explicitly signal category types such as "causes", "an effect," "remedy" and their synonyms were not used. However, at times this restriction posed a challenge to preserving the readability. Among 22 sentences identified as "cause", only once was "caused" used and twice was "lead to" used. In 26 sentences identified as "effect", the word "result" was used twice, "causes" once, and "repercussions" once.

Each sentence was divided into idea units yielding 223 idea units for the full text. I adopted Dunlosky, Hartwig, Tawson and Lipko's (2011) operational definition of an idea unit, i.e., an intermediate conceptual unit of information neither an atomic proposition (smallest unit of information, e.g. "blue", "car") nor a set of complex propositions involving multiple atomic propositions in a sentence". As seen in the example below, the sentence is categorized as a remedy and is divided into three idea units:

> Once the problem is identified, every attempt should be made to stop the release of oil (ID176), either by closing valves (ID177) or deflecting the oil into a containment area (ID178) (remedy).

My coding of targets in the text was corroborated by three graduate students who individually read the text without codes and coded each sentence as a representation of one of the four categories. Discrepant codes were discussed until complete consensus was reached. Then each idea unit was identified as a typed target in nStudy as a description, cause, effect, or remedy following the category of the sentence it is part of.

These targets are invisible to participants. nStudy's target feature allowed the software to register how participants' marking of text related to each category of information (See Appendix A). A target was considered marked if learners marked all of or any part of an idea unit.

3.3. Measures

A demographic questionnaire asked participants to identify their sex, major and age. Participants listed in bullet form seven to eight important ideas they knew about oil spills, and causes of, effects and remedies for oil spills. Each valid bullet point was awarded a score of 1 in one of two categories: prior knowledge related to ideas in the reading text or general prior knowledge. When scoring remedies, bullet points expressing prevention were not given credit because a remedy is a response to a problem. A problem has to happen first, and then a remedy follows. Prior knowledge scores were computed by summing points in each category.

Because the research focused on the metacognitive component of text marking, I selected *Metacognitive Self Regulation*, *Critical Thinking*, *Rehearsal* and *Elaboration* subscales from the *Cognitive and Metacognitive strategies* component of the *Motivated Strategies for Learning Questionnaires* (MSLQ; Pintrich, Smith, Garcia & McKeachie, 1991). Throughout, the term "reading" was changed to "studying" because studying implies using study techniques that vary from reading. Two further adaptations were made. The first adaptation requested participants to respond in light of how they generally study for "their courses", referred to here as "General MSLQ". For example, the prompt "*When reading for this course*, I make up questions to help focus my read" was changed to "*When studying,* I make up questions to help focus my studying". Participants filled out general MSLQ questionnaire before studying the reading text. The second adaptation, "MSLQ specific" requested participants to answer the questions in reference to how they studied the "oil spills text". For instance, "*When reading for this course*, I make up questions to help focus my read" was adapted to "*When I studied oil spills*, I made up questions to help focus my read". "MSLQ specific" was an interpolated task after studying the text and before answering the achievement questions. (See Appendix D)

Two measures of achievement were developed for posttests. A block of 13 transfer questions required participants to use information they learnt in different contexts relating to the description, causes, effects and remedies for oil spills (See Appendix B). Testing transfer was rarely observed in studies examining text marking. A block of 13 prompted recall questions followed transfer items to avoid presenting information upon which transfer might be based (See Appendix B). Prompted recall was selected over free recall because I wanted to pose questions about specific information presented in the text. These latter items allowed examining effects of marking on incidental learning. The test questions were distributed across categories of information as shown in Table 3.1. ensuring adequate coverage of information participants' study.

Table 3.1 Number of Questions Targeting Each Category

Text Categories	Prompted Recall	Transfer
Description	3	2
Cause	3	3
Effect	2	4
Remedy	5	4
Total	13	13

Two scoring rubrics, one for transfer questions and another for prompted recall were developed. The rubric for transfer items included correct answers, a break down of points for each answer, and idea units in the text that were sources of information needed to answer each item (See Appendix B). For each prompted recall question, every idea unit in the text needed to fully answer the question was identified, and a score of 1 was credited for each idea unit applied or recalled. Idea units with the same meaning were treated as one idea unit. Synonyms for words were accepted when marking, for example, biological agents and bio-accelerants.

To ensure objective scoring of transfer and prompted recall achievement items, a co-rater and I first collaboratively scored responses to achievement items on a randomly selected sample of two participants' data, then independently scored 8 randomly drawn participants' responses using the scoring rubrics for prompted recall and transfer. Cohen's Kappa was 0.84 for prompted recall and 0.86 for transfer.

3.4. Research Design

The research design was not balanced. Participants were randomly assigned to one of three groups varying on two independent variables: (1) criteria provided to focus study and (2) instructions to mark. Focus for study had two levels: learners were instructed to focus on (1) remedies of oil spills or (2) whatever they perceived as important. This contrast is referred to as "criteria provided and marked" vs. "free marking".

Instructions to mark had three levels: (a) mark remedies, (b) mark whatever is perceived as important, or (c) no marking. Contrasts regarding instructions are referred to as "criteria provided and marked" vs. " free marking" groups, and "criteria provided and marked" vs. "criteria provided and did not mark".

Effects of these independent variables were assessed by a between-subject design. In some analyses, I also examined within-subjects effects.

Instructions given to the **Criteria provided and marked** group (for brevity, is referred to as criteria and marking) were: "Study this text. You will be tested about remedies of oil spills. Please highlight only the remedies as you study. Carefully select what you highlight because you won't be able to erase highlights"

Instructions given to the **Criteria provided and did not mark** group (is referred to as criteria and no marking) were: *"*Study this text. You will be tested about remedies of oil spills".

Both groups that were provided criteria were given the same specific learning objective: "You will be tested about remedies of oil spills".

The **Free-marking** group was instructed: "Study this text. You will be tested about the text. Highlight as you normally do when you study for a test. Carefully select what you highlight because you won't be able to erase highlights." Participants in this group also listed standards they used for marking text after they answered the achievement questions. They were shown the second section of the text with highlights they made and were asked: "Review your highlights. Why did you mark each of these parts?" Marking and performance of this group provides a comparison to the two groups with specific study objectives.

At this point, it is important to differentiate between two terms that are used through out the paper; instructions given to participants provided "criteria" with the intention that participants would adopt them as "standards". "Criteria" is provided to participants, while "standards" are created by participants. "Standards" is the construct and "criteria" is the operational definition.

3.5. Procedure

Participants were welcomed to the lab and given the consent letter to sign. Then they were requested to choose without replacement one card from a set of three to randomly assign themselves to a group. After the third participant was assigned to a group, the cards were replaced for the next cycle of random assignment of participants to groups.

A short training session followed. It consisted of a PowerPoint presentation supplemented with a 102 word paragraph about hypertension, its description, causes, effects and remedies. First I briefly stated that when a problem is introduced, a description of the problem, plus examples of causes, effects and remedies are usually provided. Participants were shown the paragraph, then one at a time: description was bolded in red with the label "description" attached to it, followed by cause in bold red and the label "cause", then, effect and the label " effect" and finally remedy with the label "remedy". All three groups received this training to standardize expectations about the primary study task.

Then participants in the two marking groups were escorted to another room and told that to highlight text using nStudy software, they just needed to select it by dragging the cursor across text. They were also informed that once they highlighted text, they could not erase the highlight.

After training, participants sat at a computer where they accessed a bookmark matching the card chosen earlier representing their assignment to group. There were three bookmarks, each accessing one of three surveys created and presented using the online service FluidSurvey. Each survey included the same questionnaires about prior knowledge and demographics, the same text to study, and the same transfer and prompted recall achievement questions. A "***Next***" button at the bottom of each webpage allowed participants to progress to the following page. To prevent reviewing, participants could not retreat to any previous page.

Participants studied as long as they desired. When learners finished studying, they took the achievement tests. The study session took approximately an hour. After participants finished both posttests, they were compensated $12 for their time. See Appendix C for work flow on Fluidsurvey.

Chapter 4. RESULTS

This research examined the effects of standards learners use to metacognitively monitor interaction with text while reading and marking. The main questions investigated were: (1) Is there a difference in performance on prompted recall and transfer tasks when learners are provided criteria describing what to learn and mark text as they study vs. do not mark text? (2) Does providing criteria describing what to learn affect learners' performance on recall and transfer tasks targeting information categories specified or not specified in the given criteria? (3) Does marking affect learners' performance on recall and transfer tasks? (4) Is marking more effective when learners are provided criteria describing what to mark vs. when they freely mark using their own standards? (5) What standards do learners report using when they are not instructed to mark particular kinds of information and they mark freely?

Participants' transfer and prompted recall performance scores on questions targeting four information categories – description, causes, effects and remedies – were first calculated then converted to a percent of total before analyses. This normalized different scale lengths described in chapter three. All scores on transfer and prompted recall used in analysis were percent of total. Data about performance were examined for normality of distributions and outliers. None of the variables was non-normally distributed; all skewness and kurtosis values were $\leq \pm 3$. However, some outliers were detected on outcome variables as shown in Table 4.1. Those cases were retained to maximize sample size and because other data for these cases was not atypical.

Table 4.1 Outliers Identified in Outcome Variables across Three Groups

	Groups					
Variable	*Criteria & marking*		*Criteria & no marking*		*Free marking*	
	No.	*Score threshold*	*No.*	*Score threshold*	*No.*	*Score threshold*
Transfer causes	2	≥ 50	3	≥ 50	-	-
Transfer effects	3	≥ 46 (2), 10> (1)	-	-	-	-
Transfer remedies	1	>70	-	-	-	-
Total transfer categories	1	=50	-	-	-	-
P. recall description	1	=50	1	=35	1	=57
P. recall effects	-	-	-	-	1	=100
P. recall remedies	-	-	3	≥ 45 (2), =0 (1)	-	-
Total p. recall categories	-	-	-	-	1	=62

Cronbach alpha reliability coefficients were *0.672* for transfer items and *0.790* for prompted recall test items. Transfer reliability is less internally consistent because transfer questions by nature require students to apply knowledge to different situations; transfer is more divergent (less internally consistent, in the psychometric sense) than recall.

The total prompted recall variable was created by summing the percent scores on all four categories of information then dividing by four. This ensured each of the four categories was equally weighted. The same was done with total transfer variable.

Treatment fidelity was tested. First, a variable representing percent of total marking of information in categories not identified by instructions (i.e., description, causes, effects) was computed in the criteria and marking group. Then, a frequentist test (Bradley & Farnworth, 2013) was used to test whether the observed proportion of marks applied to information not specified in instructions differed from 0. Zero was the expected proportion because learners were instructed to only mark remedies. Since learners did not have the option to delete markings they made by mistake, a mistakes (false positive) rate was set at 0.10. The critical value of “a binomial random variable with parameters sample size and false positive” (Bradley & Farnworth, 2013, p. 258) was 7.876. I surveyed the Z-scores calculated for each of the 22 learner’s marking of non-remedy categories. Only two did not conform to instructions to mark remedies; their Z-scores of 28.27 and 10.64 exceeded the critical value. Z-scores for the remaining 20 learners ranged between -2.73 and 3.96. I judged the treatment was implemented with fidelity based on a 20 of 22 participants marking predominantly remedies.

The following analyses examine the contrasts referred to in the research design section in chapter 3.

4.1. Criteria and Marking And Criteria and No Marking Groups

Data from the free marking group was filtered out to focus analysis only on the criteria and marking and the criteria and no marking groups.

Since both groups were requested to focus on remedies while studying, scores on items not identified by the criteria, i.e., descriptions, causes and effects, were summed for prompted recall, then transformed into a percentage of that composite. The Cronbach alpha reliability coefficient of 0.747 for merged prompted recall items supported using this merged variable. The merged variable is now a longer test thus likely more reliable. For performance on transfer, the internal consistency for the merged variable was 0.388. Thus, a decision was made to perform separate analyses involving transfer tasks for each category.

Correlations among measures of merged prompted recall, prompted recall on remedies, transfer on description, transfer on causes, transfer on effects, transfer on remedies are reported in Table 4.2. Given generally weak correlations, a decision was made to run separate analyses for prompted recall tasks and transfer tasks.

4.1.1. Prompted Recall Analyses

At first, I decided to use scores for related prior knowledge as a covariate in the analysis examining differences between the criteria and marking and the criteria and no marking on prompted recall items. Related prior knowledge for four categories was selected over general prior knowledge because it is more related to information in the text. Related prior knowledge scores for four categories were summed to form one composite variable, total prior knowledge. Having more items generally increases reliability of the measure, a desirable feature for a covariate, especially given that the way prior knowledge was measured in this study does not allow calculating internal consistency for the measure. A further requirement for using a variable as a covariate is correlation with the outcome variable. A correlation of 0.3 was set as the threshold for identifying a covariate.

Table 4.2 Correlations of Prompted Recall Remedies, Prompted Recall Merged Categories, Transfer Description, Transfer Causes, Transfer Effects and Transfer Remedies

	P. Recall remedies	*P. Recall merged*	*Transfer descript.*	*Transfer causes*	*Transfer effects*	*Transfer remedies*	*Total prior Know.*
P. Recall remedies	1	0.192	0.286	0.076	0.449**	0.567**	0.204
P. Recall merged		1	0.526**	0.09	0.172	0.258	0.411**
Transfer descript.			1	0.039	0.308*	0.213	0.285
Transfer causes				1	0.043	-0.029	0.053
Transfer effects					1	0.386**	0.018
Transfer remedies						1	0.247
Total prior Know.							1

Note: ** p<0.01; *p<0.05

After examining correlations (Table 4.2) among total prior knowledge and prompted recall remedies and merged prompted recall, a decision was made to use total prior knowledge as a covariate for merged promoted recall but not for prompted recall on remedies. An ANCOVA was computed with groups as the independent variable, the dependent variable was merged prompted recall. Total prior knowledge was the covariate. Data conformed to the assumptions of linearity of regression; the η^2 was 0.277 and R^2 was 0.169. However homogeneity of regression was violated, p= 0.030. Thus, an ANOVA was computed. No statistically detected difference was observed between groups for performance on merged prompted recall (F= 0.065, p=0.800), criteria and marking: M= 17.448, SD=11.247, n= 22, criteria and no marking: M=16.715, SD=7.491, n=22. Raw scores for performance on merged prompted recall (out of 31), criteria and marking: M=*5.40*, SD=*3.4*, n= 22, criteria and no marking: M=5.18, SD=*2.32*, n=22.

For the outcome variable of prompted recall of remedies, an ANOVA was calculated with groups as an independent variable. No statistically detected difference was observed between groups (F=2.389, p=0.130), criteria and marking: M= 32.231, SD= 19.783, n=22, criteria and no marking: M=24.380, SD= 13.274, n=22. Raw scores for performance on prompted recall of remedies (out of 11), criteria and marking: M=*3.54*, SD=*2.17*, n= 22, criteria and no marking: M=2.68, SD=*1.46*, n=22.

It is important to note that a MANOVA was not employed with merged prompted recall and prompted recall on remedies as dependent variables due to the weak correlation between both variables reported in Table 4.2.

Since no difference between groups was detected in performance on prompted recall tasks, within-groups analysis was conducted to compare performance on remedies, for which criteria were provided and other categories of information (description, causes, effects) for which criteria were not provided. A paired-samples *t*-test showed a statistically detectable difference favouring prompted recall on remedies (M=28.30, SD=17.11) compared to prompted recall of the merged variable (M=17.08, SD=9.45); t= -4.160, p<0.000. Learners recalled more idea units identified by the provided criteria (remedies) than categories not mentioned in criteria (description, causes and effects).

Addressing the argument that remedies maybe more memorable than other categories of information in the reading text, a paired *t*-test was computed to compare differences in performance on the prompted recall merged variable (description, causes, effects) and prompted recall on remedies for the free marking group. The paired-sample t-test showed statistically detectable differences in performance for prompted recall on merged variable (M=22.43, SD=14.30) versus prompted recall on remedies (M=34.29, SD=22.78); t=(-3.58), p>0.02. This places a limitation on the interpretation of learners recalling more idea units pertaining to criteria provided in the reading instructions (remedies).

4.1.2. Transfer Analyses

For transfer tasks, a decision was made not to use total prior knowledge as a covariate following generally weak correlations reported in Table 4.2 among total prior knowledge and transfer on four categories. Because there is a possibility there could be meaningful correlations between separate categories of prior knowledge and transfer categories, correlations among each category of prior knowledge and each transfer category were calculated. As shown in Table 4.3, there were only two statistically detectable correlations. Prior knowledge, either as a sum or as separate categories, was judged not to be a useful covariate.

Table 4.3 Correlations of Prior Knowledge Description, Prior Knowledge Causes, Prior Knowledge Effects, Prior Knowledge Remedies, Transfer Description, Transfer Causes, Transfer Effects and Transfer Remedies

	Prior descript.	*Prior causes*	*Prior effects*	*Prior remedies*	*T description*	*T causes*	*T effects*	*T remedies*
Prior descript.	1	0.488**	0.351**	0.399**	0.176	0.329*	-0.020	0.211
Prior causes		1	0.112	0.613**	0.346*	-0.014	0.019	0.346*
Prior effects			1	0.195	0.063	-0.100	-0.006	-0.121
Prior remedies				1	0.201	-0.022	0.075	0.288
T description					1	0.039	0.308*	0.213
T causes						1	0.043	-0.029
T effects							1	0.386**
T remedies								1

Note: ** p<0.01; *p < 0.05

On the other hand, the prompted knowledge measure seemed a promising covariate to use for two reasons: (1) it reflects knowledge participants gained by studying the text which might apply in answering transfer items, and (2) it was statistically detectably correlated with transfer performance for three categories of information description, effects and remedies (Table 4.4). Therefore, total prompted recall was used as a covariate in a MANCOVA computed to examine transfer performance on four categories by the criteria and marking and criteria and no marking groups. I acknowledge this analysis may be underpowered because transfer causes did not correlate with total prompted recall.

Table 4.4 Correlations of Total Prompted Recall, Transfer Description, Transfer Causes, Transfer Effects and Transfer Remedies

	Total prompted recall	*T. description*	*T. causes*	*T. effects*	*T. remedies*
Total prompted recall	1	0.543**	0.194	0.338*	0.381*
Transfer descript.		1	0.039	0.308*	0.213
Transfer causes			1	0.043	-0.029
Transfer effects				1	0.386**
Transfer remedies					1

Note: ** p<0.01; *p<0.05

Groups was the independent variable. Dependent variables were transfer on description, transfer on causes, transfer on effects, and transfer on remedies. Data conformed to the assumptions of homogeneity of regression. The covariate by outcome interaction was not statistically detected by a multivariate test, $p = 0.792$.

Box's test of equality of the variance-covariance matrices was not statistically detected-indicating there were no statistically detectable differences between the covariance matrices (Meyers, Gamast & Guarino, 2013). Because assumptions were not violated, Wilks' Λ was employed to evaluate the multivariate effect. A difference was not statistically detected (*F*= 0.132, *p*=0.970). Means and standard deviations for both groups are presented in Table 4.5.

Table 4.5 Means and Standard Deviations for Transfer on Descriptions, Transfer on Causes, Transfer on Effects and Transfer on Remedies

	Criteria & Marking (n=22)		Criteria & No Marking (n=22)	
	M	*SD*	*M*	*SD*
Transfer descriptions	59.09 *(1.18)*	42.64 *(0.85)*	65.90 *(1.31)*	38.99 *(0.77)*
Transfer causes	26.13 *(2.09)*	12.74 *(1.01)*	25.00 *(2.00)*	14.94 *(1.19)*
Transfer effects	28.32 *(3.68)*	10.97 *(1.42)*	30.41 *(3.95)*	14.72 *(1.91)*
Transfer remedies	20.00 *(3.00)*	16.32 *(2.44)*	20.00 *(3.00)*	12.51 *(1.87)*

*Note: Means and Standard deviations of raw scores are provided in brackets. Maximum scores were: transfer descriptions 2, causes 8, effects 13, and remedies 15.

Since differences were not detected between groups, a within-group analysis was conducted to compare performance of information noted by the criteria and information not noted. For three paired *t*-tests computed, the *p* value was divided by four, a simple Bonferonni correction acknowledging the earlier t-test calculated for the prompted recall analysis. Thus, an adjusted alpha level of 0.0125 was used to statistically test for differences. Paired-samples *t-tests* showed statistically detectable differences for performance for transfer on description (*M*=62.50, *SD*=40.52) versus transfer on remedies (*M*=20.00, *SD*=14.37); *t*=7.045, *p*<0.001; and transfer on effects (*M*=29.37, *SD*=12.879) versus transfer on remedies (*M*=20.00, *SD*=14.37), *t*=4.103, *p*<0.001. On transfer tasks, learners performed better on questions targeting categories of information not specified in the criteria (description and effects) than on questions pertaining to categories specified (remedies).

4.2. Criteria and Marking And Free Marking Groups

4.2.1. Marking Activity

Participants' marking for each category was summed then converted into percent before analyses to normalize different scale lengths given the different number of marked text extracts pertaining to each category. I decided to use each of the four categories separately in the following analysis to provide a clearer picture of differences in marking behaviour between both groups.

First, marking activity in the two groups (criteria and marking and free marking groups) was examined for normality of distributions and outliers. In the free marking group, none of the variables was non-normally distributed; all skewness and kurtosis values were ≤ ±3. There was one outlier identified for marking remedies. However, in the criteria and marking group, kurtosis was 9.76 for marking descriptions, 6.017 for marking causes and 17.579 for marking effects. Further analysis identified two outliers. When these outliers were removed and analyses rerun, three more outliers were detected. A decision was made to retain all data to maximize sample size and apply a non-parametric test, the Mann-Whitney *U*, to examine differences between the two groups' marking behaviour of descriptions, causes and effects.

Table 4.6 Medians and Mean Ranks for Marking Descriptions, Causes and Effects for Both Marking Groups

	Descriptions		*Causes*		*Effects*	
	Median	*Mean rank*	*Median*	*Mean rank*	*Median*	*Mean rank*
Criteria & marking (n=22)	6.25	13.16	0	13.14	0	12.64
Free marking (n=22)	46.875	31.84	43.478	31.86	58.928	32.36

Medians and mean ranks for both groups are shown in Table 4.6. A statistically detectable effect of group was observed; for marking descriptions; Whitney U= 36.500, Z=-4.871, $p<0.01$; for marking causes; U=36.000, Z= -4.959, p<0.01; and for marking effects; U=25.000, Z= -5.204, p< 0.01. The free marking group marked more descriptions, causes and effects than the group provided criteria to guide marking of remedies.

Since the marking of remedies variable met the assumptions of normality, between-subjects ANOVA was computed to compare marking of remedies between the criteria and marking group (M=52.272, SD=21.054, n=22) and the free marking group (M= 44.580, SD=23.035, n=22). No statistically detectable difference was observed (F=1.33, p=0.254).

4.2.2. Performance Differences between Marking Groups

For both groups that marked text, correlations were calculated among measures of transfer on description, transfer on causes, transfer on effects, transfer on remedies, prompted recall on description, prompted recall on causes, prompted recall on effects and prompted recall on remedies (Table 4.7).

Table 4.7 Correlations of Transfer and Prompted Recall Measures for Each Category

	T. descript.	*T. causes*	*T. effects*	*T. remedies*	*P. Recall descript.*	*P. Recall causes*	*P. Recall effects*	*P. Recall remedies*
T. description	1	-0.022	0.318*	0.169	0.259	0.324*	0.326*	0.265
T. causes		1	0.305*	0.363*	0.261	0.264	0.440**	0.314*
T. effects			1	0.421**	0.241	0.288	0.497**	0.663**
T. remedies				1	0.230	0.409**	0.478**	0.597**
P. Recall descript.					1	0.473**	0.430**	0.223
P. Recall causes						1	0.560**	0.420**
P. Recall effects							1	0.656**
P. Recall remedies								1

Note: ** p<0.01; *p<0.05

Given moderate correlations among outcomes, a MANOVA was computed with Groups as an independent variable and transfer on description, transfer on causes, transfer on effects, transfer on remedies, prompted recall on description, prompted recall on causes, prompted recall on effects and prompted recall on remedies as dependent variables. Box's test of equality of the variance-covariance matrices was not statistically detected indicating there are no statistically detectable differences between the covariance matrices (Meyers et al, 2013). Therefore, Wilks' Λ was employed to evaluate all multivariate effects. Bartlett's test of sphericity was statistically detected indicating sufficient correlation among the dependent variables. A multivariate difference was not statistically detected using Wilks' Λ (F= 0.741, p=0.655). Means and standard deviations for both groups are reported in Table 4.8.

Table 4.8 Means and Standard Deviations for Performance of Tranfer and Prompted Recall on Four Categories of Information for Marking Groups

	Criteria & Marking (n=22)		Free Marking (n=22)	
	M	*SD*	*M*	*SD*
Transfer descriptions	59.09 *(1.18)*	42.64 *(0.85)*	68.18 *(1.36)*	32.89 *(0.65)*
Transfer causes	26.13 *(2.09)*	12.74 *(1.01)*	27.84 *(2.22)*	18.47 *(1.47)*
Transfer effects	28.32 *(3.68)*	10.97 *(1.42)*	27.97 *(3.63)*	17.04 *(2.21)*
Transfer remedies	20.00 *(3.00)*	16.32 *(2.44)*	21.51 *(3.22)*	19.4 *(2.91)*
PR on description	14.61 *(2.04)*	14.02 *(1.96)*	16.23 *(2.27)*	13.44 *(1.88)*
PR on causes	17.83 *(2.31)*	12.86 *(1.67)*	23.77 *(3.09)*	17.9 *(2.32)*
PR on effects	26.13 *(1.04)*	19.63 *(0.78)*	39.77 *(1.59)*	26.34 *(1.05)*
PR on remedies	32.23 *(3.54)*	19.78 *(2.17)*	34.29 *(3.77)*	22.78 *(2.50)*

*Note: Means and Standard deviations of raw scores are provided in brackets. Maximum scores were: transfer descriptions 2, causes 8, effects 13, and remedies 15. Maximum scores for prompted recall were: descriptions 14, causes 13, effects 4, and remedies 11.

4.2.3. Text Marking and Recall

The criteria and marking group and the free marking group were compared on recall of idea units marked and idea units not marked. If an idea unit was marked and recalled, it was given a score of 1. These scores were added to create a total of the number of idea units marked and recalled. That sum was divided by the number of idea units marked (multiplied by 100) to form a score describing whether marked content was recalled (R_M). A recalled not-marked variable was formed by dividing the number of idea units not marked that were recalled by the number of not-marked idea units multiplied by 100 (R_NOT_M). Then a MANOVA was calculated with groups as independent variable and *R_M* and *R_NOT_M* as dependent variables. Box's test of equality of the variance-covariance matrices was not statistically detected (Meyers et al, 2013). Therefore, the assumption is not violated and Wilks' Λ was employed to evaluate all multivariate effects. Bartlett's test of sphericity was statistically detected. There were no statistically detected differences between groups according to Wilks' Λ value (F= 0.355, p=0.703) means and standard deviations for both groups are provided in Table 4.9. Therefore, a within-groups analysis was conducted to compare R_M and R_NOT_M. A paired-samples t-test showed a statistically detectable difference for R_M (M=32.408, SD=20.153) and R_NOT_M (M= 16.588, SD=10.573); t=4.926, p<0.001. The conditional probability of recalling marked content was approximately two times greater than non-marked content.

Table 4.9 Means and Standard Deviations for Idea Units Recalled and Marked, and Recalled and Not Marked for Marking Groups

	Criteria & Marking (n=22)		Free Marking (n=21)	
	M	*SD*	*M*	*SD*
Recalled & marked	30.01	17.51	34.91	22.76
Recalled & not marked	15.88	9.4	17.32	11.86

4.3. MSLQ, Text Marking and Performance

The twenty-two MSLQ general items yielded a Cronbach alpha reliability coefficient of 0.783; Cronbach alpha for the MSLQ specific items was 0.848. A statistically detected correlation was found between MSLQ-G and MSLQ-S, r=0.330, p<0.05. A decrease in means of ratings is noted in the MSLQ specific for each item from its corresponding item in MSLQ general (Table 4.10). A paired-samples t-test showed a statistically detectable difference MSLQ-G (M= 101.242, SD= 14.677) and MSLQ-S (M=77.439, SD=18.625); t=11.323, p<0.01. When learners are asked to report how they study in general, they tend to overestimate what they report doing compared to when asked about a more specific studying event.

Table 4.10 Means on MSLQ for Criteria & Marking, Criteria & No Marking and Free Marking Groups

	Criteria & Marking Group		*Criteria & No Marking Group*		*Free Marking Group*	
	MSLQ-G	*MSLQ-S*	*MSLQ-G*	*MSLQ-S*	*MSLQ-G*	*MSLQ-S*
Item 1	3.591	2.545	4.045	3.590	3.727	3.045
Item 2	3.773	2.591	3.727	2.773	4.500	2.182
Item 3	5.727	4.364	6.136	5.000	6.091	5.091
Item 4	4.045	2.864	4.364	3.227	4.500	3.091
Item 5	3.727	3.409	4.455	3.545	4.591	2.409
Item 6	5.045	2.909	4.409	3.364	4.818	2.955
Item 7	3.955	3.091	4.091	3.455	3.591	2.727
Item 8	3.545	2.455	4.364	2.545	4.273	2.591
Item 9	4.182	4.136	4.500	5.273	4.318	4.227
Item 10	5.455	2.955	5.864	3.727	4.955	3.955
Item 11	4.545	3.318	4.773	3.136	4.909	3.000
Item 12	4.409	3.136	4.409	3.727	4.591	2.591
Item 13	4.136	3.409	5.091	4.136	4.000	3.045
Item 14	4.000	2.864	4.136	3.727	4.364	2.955
Item 15	4.364	3.545	5.682	3.636	4.364	3.909
Item 16	3.818	3.409	4.318	3.227	3.455	2.500
Item 17	5.000	3.136	5.227	3.545	4.776	3.182
Item 18	5.682	3.682	5.682	5.227	5.727	4.909
Item 19	4.045	2.818	3.864	2.636	3.818	3.545
Item 20	4.591	4.136	5.727	3.818	4.727	3.364
Item 21	5.045	4.909	5.727	5.773	5.364	5.591
Item 22	4.318	4.318	5.500	4.545	4.591	3.818

Focusing analysis on the criteria and marking and criteria and no marking groups, correlations were calculated between performance on prompted recall (percentage of total performance on promoted recall questions on descriptions, causes, effects and remedies), performance on transfer (percentage of total performance on transfer questions on descriptions, causes, effects and remedies), and learners' ratings on MSLQ general and MSLQ specific. No statistically detectable correlations were observed. (Appendix D)

Focusing analysis on the criteria and marking group and the free marking group, correlations were computed for performance on transfer and prompted recall and ratings on MSLQ general and specific, again no statistically detectable relations were noted (Appendix D). Again, there were no statistically detectable correlations between learners' marking and ratings on MSLQ general and specific. (Appendix D)

4.4. Marking Criteria for Free Marking Group

Table 4.11 shows reasons for marking (metacognitive standards) provided by participants in the free marking group (n=22) and the frequency each was mentioned. Some reasons were merged, for example “Statistics”, “figures” and “number” were collapsed, as were “key points” and “important.” It should be noted that 20 learners provided more than one standard. For instance, one learner mentioned marking “causes, effects, remedies and important” information.

Table 4.11 Standards for marking and Frequency Provided by the Free Marking Group

	Standards	*Frequency*
1	Important	18
2	Causes	10
3	Effects	9
4	Remedies	9
5	Definitions	6
6	Comparisons	6
7	Stat/numbers/figures	6
8	Description	6
9	Things did not know about	5
10	Interesting	4
11	Terms	3
12	Relevant	3
13	Contrary to personal belief	1
14	Gist of information	1
15	Topic sentence	1
16	Adjective describing important terms	1
17	Things not clear about	1

The top four reasons for marking text were: important, causes, effects and remedies. Important was the most commonly reported standard for marking; 18 out of 22 participants reported using it to guide their marking.

Chapter 5. DISCUSSION

This research was designed to answer the following questions: (1) Is there a difference in performance on prompted recall and transfer tasks when learners are provided criteria describing what to learn and mark text as they study vs. do not mark text? (2) Does providing criteria describing what to learn affect learners' performance on recall and transfer tasks targeting information categories specified or not specified in the given criteria? (3) Does marking affect learners' performance on recall and transfer tasks? (4) Is marking more effective when learners are provided criteria describing what to mark vs. when they freely mark using their own standards? (5) What standards do learners report using when they are not instructed to mark particular kinds of information and they mark freely?

In this chapter, I review findings related to each question, draw implications and note limitations of the study.

5.1. Is there a difference in performance on prompted recall and transfer tasks when learners are provided criteria describing what to learn and mark text as they study and when provided criteria but do not mark text?

No detectable differences were observed on transfer and prompted recall tasks between the group provided criteria that marked text and the group provided criteria that did not mark. This suggests if learners are given specific criteria to focus their learning, they do not need to mark text to productively process targeted content. As proposed in the text marking model in chapter two, learners use metacognitive standards identified by criteria provided before reading to judge whether a text element is relevant, If an element is relevant to their standards, then it gets marked. If learners are not required to mark, they still use these standards to metacognitively judge whether a text segment is worthy of cognitive processing.

Most text marking research reporting elevated performance due to marking text ignored the impact of reading objectives on learners' study. Instructions these studies provided for non-marking groups did not prompt learners to metacognitively monitor any particular information while reading, e.g., instructions were simply "*Read the text*" (Amer, 1994; Hayati & Shariatifar, 2009). In contrast, instructions given to non-marking groups in studies reporting no differences in performance between marking and non-marking groups included an objective that induced metacognitive monitoring while studying; e.g. "*Read the passage and re-read any section not comprehended*" (Jonassen, 1984). My findings suggest learners do not have to physically mark text. Prompting learners to engage in metacognitive monitoring of whether content is relevant to standards, is enough. This calls attention to the judgement learners make when deciding whether a text element aligns with standards or not in the context of marking text.

It is important to note, the between-groups comparisons examines aggregate performance score across all items. It does not focus on recall or transfer of marked and unmarked information for marking group. In depth analyses of recall and transfer of marked and unmarked information is later covered in section 5.3.3.

5.2. Does providing criteria describing what to learn affect learners' performance on recall and transfer tasks targeting information categories specified or not specified in the given criteria?

On prompted recall tasks, learners recalled more idea units in the category specified by criteria (remedies of oil spills) than idea units related to categories not specified by criteria (description, effects and causes of oil spills). A meta-analysis by Klauer (1984) reported instructional objectives boost learning information relevant to the stated objectives (i.e., intentional learning) but undermine learning of irrelevant information (i.e., incidental learning). Instructional objectives govern intentions; intentions in turn guide learners' choice of what merits attention when studying (Klauer, 1984) and enhances retention. Klauer reasoned when learners are given specific instructions about content they need to focus on, they do not use up cognitive resources trying to find what is important and instead use criteria provided to help them locate information they should focus on. More cognitive resources are then available to process information. It's important to note that in the meta-analysis, no distinction was made between different types of outcome measures, i.e. recall, transfer, or inference. I accessed papers included in the meta-analysis and found that most studies used recall or prompted recall as outcomes. Some studies only specified the assessment format (e.g. multiple-choice questions) and not the outcome measure (e.g. inference, recognition).

Although my findings corroborate earlier findings of the instructional objectives literature, one should take them with a grain of salt. Further analysis of performance by free marking group on different categories of information revealed that text extracts identified as remedies were more memorable than extracts pertaining to other categories (description, causes and effects). This places limitations on interpretations regarding the effect of providing criteria on information recall. I caution against attributing the effect on recall to providing criteria alone, other variables were at play. This finding could inform design of future research, where category selected to be provided as criteria for marking should not be the most memorable.

With transfer tasks, learners performed better on questions targeting categories of information not specified in given criteria (description and effects) than on the category identified in criteria (remedies). There is a possible reason why learners performed better on transfer questions on description than on remedies. Transfer questions about descriptive information were multiple-choice questions unlike the short-answer format of transfer questions on effects and remedies. There could be a confound between the kind of knowledge description category entails, i.e. descriptive, and a method effect, i.e., the item format. However, this doesn't provide grounds for speculating why learners performed better on transfer questions about effects than about remedies. It would be important to discover whether it is a replicable effect and if it didn't then it's worth trying to figure out why it happens.

5.3. Is marking more effective when learners are provided criteria describing what to mark vs. when they freely mark using their own standards?

5.3.1. Differences in Marking Activity

Results show the group provided criteria to mark text marked less information about categories not specified in the criteria than the free marking group did. However, there was no statistically detectable difference between these groups' marking of information identified by the criteria (remedies). Providing learners with criteria to guide marking and studying dampened marking of content not specified in the criteria, but it did not elevate marking criteria-related content. One might expect learners given criteria would mark almost all content fitting the criteria, but this was not the case. They marked only about 52% of criteria-related text. That they do mark all this information suggests they could identify content pertaining to criteria, using either or both text related cues and prior knowledge. That is, they metacognitively monitored what they marked but they were no more active in marking criteria-related text than learners who marked text freely according to their own standards.

Findings indicate (a) brief training can affect learners' text marking and (b) learners used instructions provided before reading to create standards for metacognitively monitoring text to mark. But there is still more to explore about other standards used when deciding not to mark information pertaining to criteria given. Further research is needed to examine other standards learners use to decide what needs to be marked.

5.3.2. Differences in Performance

No differences in performance on prompted recall or transfer tasks were detected between learners given criteria about what to mark and learners who freely marked. To interpret this finding I refer to instructions given to free marking group and draw on other findings reported in this study. The free marking group was instructed to "*Study this text. You will be tested about the text. Highlight as you normally do when you study for a test. Carefully select what you highlight because you won't be able to erase highlights.*" This group reported using a variety of standards when reading and marking. Standards they used included "important"," causes", "did not know about", and interestingly, "relevant" (for an inclusive list refer to section 4.4 in Chapter 4). This finding shows that although free marking group was not given specific criteria of what to mark, they had their own standards that they used to monitor their marking.

Providing specific instructions of what to mark reduced the marking of information not specified by criteria but did not boost marking of criteria-related information. Perhaps, the criteria and marking group only marked what they perceived as "unfamiliar remedies" or "interesting remedies". They may have used other standards along with remedies to decide what to mark.

These findings imply that both the criteria and marking and free marking groups were engaged in judging whether a text extract fits their standards and warrants marking. No differences in performance between both marking groups suggest learners marking text in preparation for a test, even if they are not given specific criteria of what to mark, create their own standards to monitor their marking. It does not matter whether learners are given specific criteria to guide their marking or not as long as they are metacognitively engaged in deciding what to mark.

5.3.3. Does marking enhance learners' recall of information marked?

Recall of both marked and unmarked content was low; approximately 48% of idea units that could be recalled were actually recalled in answer to questions. The conditional probability of recalling marked content was approximately twice that of non-marked content. This finding aligns with earlier findings. Winne et al (2015) reported the probability of recalling an idea increased 50% when it was marked. Another study revealed the probability of recalling a marked idea unit was 1.96 times more than recalling a non-marked idea unit (Winne et al., 2017). Fowler and Barker (1974) showed marked information was better retained than non-marked; so did Nist and Hogrebe (1987). Also, Blanchard and Mikkleson (1987) found marked information had 50% greater chance of being recalled than non-marked, regardless of learners' reading ability and study time.

All these findings support the selective attention model proposed by Anderson and colleagues (1982) where noting, operationalized as marking in these studies, is hypothesized as crucial to learning. According to that model, content is judged for importance while being processed, then information identified as important is noted, e.g. marked or written in notes. Noting is theorized to enhance attention given to noted content, thus noted text is better learned.

As mentioned earlier in chapter 2, not all text-marking research that analysed learners' marking supported the marked-recalled relationship. Some studies introduced variables in-addition to marking text which might affect recall, e.g., level of importance of information in text. These studies reported important information was learnt regardless of whether it was marked. For instance, Johnson (1988) defined superordinate sentences as "sentences, which introduced the topic of the paragraph" and subordinate sentences as sentences that succeeded superordinate sentences (p. 22). That study reported learners in a text-marking group recalled more superordinate sentences, whether marked or not, than subordinate sentences. Wade and Trathan (1989) explored relations between importance of information, noting (e.g., marking or writing notes), and learning. They found important information was retained regardless of whether it was noted. In the current study, analyses to examine the effect of marking on recall, discussed in this section, focused on calculating the probability of recalling marked information irrespective of its category. Factoring in category of information was beyond the scope of this question.

Marking text does enhance the probability of recalling marked text. One might think this finding contradicts my very first finding that "physically marking text is not important" (Section 5.1), but this is not correct. The finding reported in section 5.1 suggests marking text is a meta-phenomenon resulting from a learner's judgement of whether a text segment is relevant to standards, thus worthy of attention. The efficacy of text marking arises from learners' metacognitive judgement about deciding what content to mark.

Johnson (1988) reported similar findings. That study found no differences between marking and read-only groups on recall aggregated across all question items. However, analysis of recall of marked and non-marked information showed elevated recall of marked information when compared to non-marked.

5.4. What standards do learners report using use when they are not instructed to mark particular kinds of information and they mark freely?

Learners reported 17 different standards for judging what to mark. The top four were: important (81%), causes (45%), effects (40%) and remedies (40%). Important is the most commonly used standard for marking. The short training that introduced learners to different categories of information in the reading text, even though not directive as in saying "Focus on X", appears to have influenced reports about what kinds of information to mark. However, this influence was not very prevailing for the following reasons: First, the four categories introduced in training were "description, causes, effects and remedies", learners marked more causes, effects and remedies than description. If training influenced metacognitive monitoring, learners would have marked more description. Second, kinds of information – description, causes, effects, and remedies – were mentioned 34 times, while other reasons for marking text were mentioned 56 times. Categories of information introduced in the training session did not overwhelm participants' other reasons for marking text.

5.5. MSLQ: Specific and General

When learners are asked to report how they study in general, they tend to overestimate what they report doing when asked about a more specific studying episode. This implies it is better to provide learners with learning questionnaires that identify a specific context for learners to consider when providing ratings to how they study.

5.6. Instructional Implications

In the context of encoding information in text, my study has the following instructional implications and recommendations:

1. Short training can affect learners' text-marking activity.
2. Instructors do not have to request learners to mark text as long as learners are prompted to engage in metacognitive monitoring to decide which content merits attention. In cases where no specific reading objectives are provided to learners, a brief description of the main categories of information of the reading text would compensate for the lack of specific reading objectives.

It is important to view these implications within the context of this research where the reading text used is approximately 1800 words.

5.7. Limitations

The "free marking" group was presented with a section of the reading text that included their marking and was requested to list reasons for marking. Four participants reported that they marked parts by mistake and could not erase marking because the software did not allow it. This could have a bearing on content identified as marked.

The study did not request the criteria and marking group to report standards, other than remedies, they used to monitor their marking. In retrospect, a fuller picture of the decision making processes that learners use could have been developed had this group also been asked to identify standards they used to mark text.

To survey prior knowledge, learners were requested to write in bullet form seven to eight important ideas they knew about oil spills, and note causes of, effects and remedies for oil spills. This measure might not be considered as reliable as providing specific questions or multiple-choice questions. However, it has the advantage of avoiding priming learners to focus on specific information when reading and studying.

Transfer questions on description were posed in multiple-choice format unlike questions on other categories that were in short answers format. Learners' performance on description questions could not be easily interpreted due to the confound between the kind of knowledge assessed the method effect, i.e. multiple-choice format.

Multiple statistical analyses of data from a single sample raises caution about probability of type I error. Even though I used a Bonferonni correction for each analysis involving multiple t-tests, the chance of type I error is inflated across the experiment-wise set of inferential statistical tests. However, I feel justified in carrying out the analysis I did because it is common practice and the threshold one sets for alpha is arbitrary. It is also important to note that low power may have masked opportunity to identify statistical differences between groups.

References

Abd-El-Fattah, S. M. (2011). The effect of test expectations on study strategies and test performance: A metacognitive perspective. *Educational Psychology, 31(4),* 497-511. http://dx.doi.org/10.1080/01443410.2011.570250

Aberšek, B. (2015). Changing educational theory and practice. *Problems of Education in the 21st Century, 66,* 4-6.

Amer, A. A. (1994). The effect of knowledge map and underlining training on the reading comprehension of scientific texts. *English for Specific Purposes, 13 (1),* 35-45. http://dx.doi.org/10.1016/0889-4906(94)90023-X

Annis, L., & Kent Davis, J. (1978). Study techniques and cognitive style: Their effect on recall and recognition. *The Journal of Educational Research, 71(3),* 175-178

Ausubel, D.P. (1968). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology, 51*, 267-272

Ayer, W. W., & Milson, J. L. (1993). The effect of notetaking and underlining on achievement in middle school life science. *Journal of Instructional Psychology, 20(2)*, 91-98

Baddeley, A. D. (1978). The trouble with levels: A re-examination of Craik and Lockhart's framework for memory research. *Psychological Review, 85(3),*139-152. http://dx.doi.org/10.1037/0033-295X.85.3.139

Baker, L. (2002). Metacognition in comprehension instruction. In C. Block & M. Pressely (Eds.), Comprehension instruction: Research-based best practices (pp.77-95). New York : Guilford.

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. Handbook of reading research, 1(353), V394.

Bell, K. & Limber, J. E. (2010). Reading skills, textbook marking, and course performance*. Literacy research and instruction, 49,* 56-67.

Bisra, K., Marzouk, Z, Guloy, S. & Winne, P. (2014). A Meta-Analysis of the Effects of Highlighting or Underlining while Studying. Poster presented at the American Educational Research Association annual conference, Philadelphia, PA.

Blanchard, J., & Mikkelson, V. (1987). Underlining performance outcomes in expository text. *The Journal of Educational Research, 80(4),* 197-201

Bradley, J. R. & Farnsworth (2013). Testing for a zero proportion. *Open Journal of Statistics, 3,* 258-260.

Braten, I. & Samuelstuen, M. S. (2006). Does the Influence of Reading Purpose on Reports of Strategic Text Processing Depend on Students' Topic Knowledge? *Journal of Educational Psychology, 96 (2),* 324–336.

Cashen, V. M., & Leicht, K. L. (1970). Role of the isolation effect in a formal educational setting. *Journal of Educational Psychology, 61(6),* 484-486

Coles, P., & Foster, J. (1975). Typographic cueing as an aid to learning from typewritten text. *Programmed Learning and Educational Technology, 12(2),* 102-108.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11(6),* 671-684. http://dx.doi.org/10.1016/S0022-5371(72)80001-X

Craik, J, M., & Martin, D. P. (1980). The effect of a presentation on how to underline a text in introductory psychology courses*. Journal of Reading, 23 (5),* 404-407

Crewe, J.C. (1969). The effect of study strategies on the retention of college text material. *Journal of Literacy Research*, 1(2), 45-52, DOI: 10.1080/10862966909546847

Crouse, J. H., & Idstein, P. (1972). Effects of encoding cues on prose learning. *Journal of Educational Psychology, 63(4),* 309-313

Di Vesta, F. J., & Gray, G. S. (1972). Listening and note taking. *Journal of educational psychology, 63(1),* 8-14

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14(1)*, 4-58. http://dx.doi.org/10.1177/1529100612453266

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational psychologist*, *34*(3), 169-189. http://dx.doi.org/10.1207/s15326985ep3403_3

Entwistle, N., & Entwistle, D. (2003). Preparing for examinations: The interplay of memorising and understanding, and the development of knowledge objects. *Higher Education Research & Development, 22(1),* 19-41. http://dx.doi.org/10.1080/0729436032000056562

Fass, W., & Schumacher, G. M. (1978). Effects of motivation, subject activity, and readability on the retention of prose materials. *Journal of educational psychology, 70(5),* 803-807

Feldt, R. C., & Ray, M. (1989). Effect of test expectancy on preferred study strategy use and test performance. *Perceptual and Motor Skills, 68(3_suppl*), 1157-1158. http://dx.doi.org/10.2466/pms.1989.68.3c.1157

Fowler, R.L. & Barker, A. S. (1974). Effectiveness of highlighting for retention*. Journal of Applied Psychology, 59(3)*, 358-364. http://dx.doi.org/10.1037/h0036750

Gier, V.S., Kreiner, D.S & Natz-Gonzalez, A. (2009). Harmful effects of preexisting inappropriate highlighting on reading comprehension and metacognitive accuracy. *The Journal of General Psychology, 136(3),* 287-300. http://dx.doi.org/10.3200/GENP.136.3.287-302

Gier, V.S., Herring, D. Hudnell, J. Montoya, J. & Kreiner, D.S. (2010). Active reading procedures for moderating the effects of poor highlighting. *Reading Psychology, 31,* 69-81. http://dx.doi.org/10.1080/02702710903397082

Glover, J. A., Zimmer, J. W., Filbeck, R. W., & Plake, B. S. (1980). Effects of training students to identify the semantic base of prose materials. *Journal of applied behavior analysis, 13(4),* 655-667

Goldstein, A. G., & Chance, J. E. (1971). Visual recognition memory for complex configurations. *Attention, Perception, & Psychophysics, 9(2),* 237-241. http://dx.doi.org/10.3758/BF03212641

Hacker, Douglas J., John Dunlosky, and Arthur C. Graesser, eds. Handbook of metacognition in education. Routledge, (2009)

Hartley, J., Bartlett, S., & Branthwaite, A. (1980). Underlining can make a difference—sometimes. *The Journal of Educational Research, 73(4),* 218-224

Hayati, A. M., & Shariatifar, S. (2009). Mapping Strategies. *Journal of College Reading and Learning, 39(2)*, 53-67. http://dx.doi.org/10.1080/10790195.2009.10850318

Hoon, P. W. (1974). Efficacy of three common study methods. *Psychological Reports, 35(3),* 1057-1058

Idstein, P., & Jenkins, J. R. (1972). Underlining versus repetitive reading. *The Journal of Educational Research, 65(7),* 321-323

Johnson, L.L. (1988). The effects of underlining textbooks sentences on passages and sentence retention. *Reading Research and Instruction, 28(1),* 18-3. http://dx.doi.org/10.1080/19388078809557955

Jonassen, D.H. (1984). Effects of generative text processing strategies on recall and retention. *Human Learning, 3,* 241-256.

Kaufman, R. (2000). *Mega planning: Practical tools for organizational success.* Thousand Oaks, CA: Sage.

Klare, G. R., Mabry, J. E., & Gustafson, L. M. (1955). The relationship of patterning (underlining) to immediate retention and to acceptability of technical material. *Journal of applied psychology, 39(1),* 40-42

Klauer, K. J. (1984). Intentional and incidental learning with instructional texts: A meta-analysis for 1970–1980. *American educational research journal, 21(2),* 323-339. http://dx.doi.org/10.3102/00028312021002323

Kobayashi, K. (2007). The influence of critical reading orientation on external strategy use during expository text reading*. Educational Psychology: An International Journal of Experimental Educational Psychology, 27(3),* 363-375. http://dx.doi.org/10.1080/01443410601104171

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14(2),* 219-224. http://dx.doi.org/10.3758/BF03194055

Kulhavy, R. W., Dyer, J. W., & Silver, L. (1975). The effects of notetaking and test expectancy on the learning of text material. *Journal of Educational research, 68*, 363-365. http://dx.doi.org/10.1080/00220671.1975.10884802

Leicht, K. L., & Cashen, V. M. (1972). Type of highlighted material and examination performance. *The Journal of Educational Research, 65(7),* 315-316

Leutner, D., Leopold, C. & Den Elzen-Rump, V. (2007). Self regulated learning with a text highlighting strategy. *Zeitschrift für Psychologie/Journal of Psychology, 215(3),* 174-182.

Linder, R., Gordon, W.I., & Harris, B. R. (1996). Highlighting text as a study strategy: beyond attentional focusing. Poster presented at the American Educational Research Association annual conference, New York, NY.

Lonka, K., Lindblom-YlÄnne, S., & Maury, S. (1994). The effect of study strategies on learning from text. *Learning and Instruction, 4(3),* 253-271

Marken, R., & Maland, J. (1979). Single-subject analyses of three study methods. *Psychological Reports, 44,* 765-766

Marzouk, Z, Winne, P., & Nesbit, J. C. (manuscript). Types of assessment, highlighting and studying.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19(2),* 113–139. http://dx.doi.org/10.1007/s10648-006-9010-7

McCrudden, M. T., & Sparks, P. C. (2014). Exploring the effect of task instructions on topic beliefs and topic belief justifications: A mixed methods study. *Contemporary Educational Psychology, 39(1)*, 1-11 http://dx.doi.org/10.1016/j.cedpsych.2013.10.001

McDaniel, M. A., Blischak, D. M., & Challis, B. (1994). The effects of test expectancy on processing and memory of prose. *Contemporary educational psychology, 19(2),* 230-248. http://dx.doi.org/10.1006/ceps.1994.1019

McKeown, M. G., Beck, I. L., & Blake, R. G. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly, 44(3),* 218-253. http://dx.doi.org/10.1598/RRQ.44.3.1

Meichenbaum, D., Burland, S., Gruson, L., & Cameron, R. (1985). Metacognitive assessment. In S. Yussen (Ed.) The growth of reflection in children (pp. 1–27). Orlando, FL: Academic.

Meyers, S. L., Gamst, G., & Guarino, A.J. (2013. *Applied Multivariate Research* (2nd ed.). Los Angeles, CA: Sage.

Nelson, T. O. (1996). Consciousness and metacognition. *American psychologist, 51(2),*102-116. http://dx.doi.org/10.1037/0003-066X.51.2.102

Nist, S. L., & Kirby, K. (1986). Teaching comprehension and study strategies through modeling and thinking aloud. *Literacy Research and Instruction, 25(4),* 254-264

Nist, S.L. & Hogrebe, M.C. (1987). The role of underlining and annotating in remembering textual information. *Reading Research and Instruction, 27(1),* 12-25. http://dx.doi.org/10.1080/19388078709557922

Nist, S. L., & Simpson, M. L. (1988). The effectiveness and efficiency of training college students to annotate and underline text. In National Reading Conference Yearbook. National Reading Conference

Pauk, W. (1974). How to study in college. Boston: Houghton Mifflin.

Peterson, S.E. (1992). The cognitive functions of underlining as a study technique. *Reading Research and Instruction, 31(2),* 49-56.

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). Ann Arbor: University of Michigan.

Rickards, J. P., & August, G. J. (1975). Generative underlining strategies in prose recall. Journal of Educational Psychology, 67(6), 860-865

Rickards, J. P., & Denner, P. R. (1979). Depressive effects of underlining and adjunct questions on children's recall of text*. Instructional Science, 8(1),* 81-90

Schnell, T. R., & Rocchio, D. J. (1974). A Study of the Relative Effectiveness of Various Underlining Strategies on Reading Comprehension. Paper presented at the Annual Meeting of the College Reading Association, Bethesda, Maryland.

Schraw, G., Wade, S. E., & Kardash, C. A. (1993). Interactive effects of text-based and task-based importance on learning from text. *Journal of Educational Psychology, 85(4),* 652-661. http://dx.doi.org/10.1037/0022-0663.85.4.652

Silvers, V.L. & Kreiner, D.S. (1997). The effects of preexisting inappropriate highlighting on reading comprehension. *Reading Research and Instruction, 36(3),* 217-223. doi: 10.1080/19388079709558240

Smith, B. D. (1985). Bridging the gap: College reading. Scott, Foresman and Company.

Company.Stordahl, K. E., & Christensen, C. M. (1956). The effect of study techniques on comprehension and retention. *The Journal of Educational Research, 49(8),* 561-570

Surber, J. R. (1992). The effect of test expectation, subject matter, and passage length on study tactics and retention. *Literacy Research and Instruction, 31(3),* 32-40

Tilstra, J., & McMaster, K. L. (2013). Cognitive processes of middle grade readers when reading expository text with an assigned goal. *Learning and Individual Differences, 28,* 66-74. http://dx.doi.org/10.1016/j.lindif.2013.09.011

Tobias, S. & Everson, H.T. (2000). Assessing metacognitive knowledge monitoring. In G. Schraw (Ed.), Issues in the measurement of metacognition (pp. 147–222). Lincoln, NE: Buros Institute of Mental Measurements and Erlbaum Associates

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. Metacognition and learning, 1(1), 3-14. http://dx.doi.org/10.1007/s11409-006-6893-0

Wade, S.E. & Trathen, W. (1989). Effect of self selected study methods on learning. *Journal of Educational Psychology, 8(1)*, 40-4. http://dx.doi.org/10.1037/0022-0663.81.1.40

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), Metacognition in educational theory and practice (pp. 277–304). Hillsdale, NJ: Lawrence Erlbaum.

Winne, P.H. (2001). Self-regulated learning viewed from models of information processing. In B.J. Zimmerman and D.H. Schunk, eds, Self-Regulated Learning and Academic Achievement: Theoretical Perspectives, pp. 153–189. 2nd edition, Mahwah, NJ: Lawrence Erlbaum Associates.

Winne, P.H., Nesbit, J.C., Ram, I., Marzouk, Z., Vystasek, J., Samadi, D., Stewart-Alonso, J., Rakovic, M., Liaqat, A., Lin, M.P.C., Mozaffari, Z., & Savoji, A. (2017). Tracing Metacognition by Highlighting and Tagging to Predict Recall and Transfer. American Educational Research Association conference, San Antonio.

Winne, P. H., Marzouk, Z., Ram, I., Nesbit, J. C., & Truscott, D. (2015, August). Effects of orienting learners to study terms or processes on highlighting and achievement. European Association for Learning and Instruction, Limassol, Cypress

Yue, C.L., Storm, B.C., Kornell, N., & Bjork, E.L. (2015). Highlighting and Its Relation to Distributed Study and Students' Metacognitive Beliefs. Educational Psychological Review, 27, 69-78, DOI 10.1007/s10648-014-9277-z

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice, 41(2)*, 64-70. http://dx.doi.org/10.1207/s15430421tip4102_2

Zimmerman, B. J. (2008). Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. *American Educational Research Journal, 45(1),* 166 –183. DOI: 10.3102/0002831207312909

Appendix A. Oil Spills Reading Text

Section 1

Oil is a substance derived from petroleum (ID1) that does not include benzene (ID 2), xylene (ID3), ethylene (ID4), toluene (ID 5) or liquefied natural (ID 6) or petroleum gas (ID 7) (description1). An oil spill is an isolated event in which oil is discharged over a relatively short time (ID 8) (description 2) due to accidents (ID 9), natural disasters (ID 10), or deliberate acts (ID 11) (cause 1).

An oil spill does not include an event in which oil leaks slowly over a long period of time (ID 12) (description 3). Nor does it include spillage during routine operations (ID 13) such as loading (ID 14) and discharging cargo (ID 15), and taking on fuel oil described in international regulations that govern discharges from oil tankers (ID 16) (description 4). Finally, spillage over a relatively long period involving less than 5 ppm (parts per million) oil discharges (ID 17) in effluents from oil refineries is not an oil spill (ID 18) even if those discharges violate pollution regulations (ID19) (description 5).

The Canadian government constantly performs routine surveillance of pipelines (ID 20) and shipping routes (ID 21) to enforce pollution laws (ID 22) and identify offenders (ID 23) (description 6). Oil spills from ruptured pipelines may go unchecked for a period of time (ID 24) when there is uncertainty of the exact location of the spill (ID 25) or limited knowledge about the extent of spillage (ID 26) (description 7). Contrary to popular perceptions about recent catastrophic events, the frequency of oil spills (ID 27) and the total volume of spillage have decreased significantly over the past 20 years (ID28) (description 8). The decline is most noticeable in the last few years (ID 29), independently of increases in social protests about oil transport (ID30) (description 9).

Section 2

The term “oil spill” is usually applied to marine oil spills (ID 31) where oil is released into the ocean (ID 32) or coastal waters (ID 33) (description 10). Spills may include refined petroleum products (ID 34) and their by-products (ID35), heavier fuels used by large ships (ID 36), such as bunker fuel (ID 37), or any oily refuse (ID38) or waste oil (ID39) (description 11). Transporting oil by ocean tankers (ID 40) and pipelines (ID41) accounts for about 70% of accidental oil spillage into the oceans (ID 42) (cause 2).

In contrast, the contribution of offshore drilling (ID43) and production activities is less than 1% (ID44) (cause 3). Marine oil spills were a major environmental problem in the 1960s (ID45) (description 12), chiefly due to intensified petroleum exploration (ID46), oil production activities over the continental shelves (ID47) and the use of supertankers capable of transporting more than 500,000 metric tons of oil (ID48) (cause 4). Governments all over the world passed laws (ID 49) and regulations to promote the sustainable growth of marine shipping without compromising safety (ID50) (description 13). Despite these laws, thousands of minor and several major oil spills related to well discharges (ID51) and tanker operations are reported each year (ID 52) (cause 5).

Marine oil spills could also originate in releases of crude oil from pipelines (ID53), offshore drilling rigs (ID54) and wells (ID55) (cause 6). These spills may result in oil pollution over large areas (ID56) and present serious environmental hazards for marine ecosystems (ID57) (effect 1). The first response to an oil spill in a marine environment is to control its spread (ID58) (remedy 1). Containment booms are used to limit spillage (ID 59) if the crew reaches the spill within an hour or two (ID60) (remedy 2). These are temporary floating barriers (ID 61) that help “box in” oil on the surface of the water (ID62) (remedy 3). The Canadyne AirBoom is one brand of pressure inflatable boom that uses air for buoyancy (ID63) (Elaboration 1-R).

Different types of oil behave differently in the environment (ID64) and animals (ID65) and birds are affected differently by different types of oil (ID66) (effect 2). For instance, there are two major types of oil (ID67); sticky heavy oil (ID68) such as bunker fuel (ID69) and non-sticky light oil (ID70) such as gasoline (ID71) (Elaboration2- E). Bunker fuel is sticky heavy oil that could be released in water through a hole in the body of the tanker (ID 72) (Cause 7). This hole could be caused by abrasion when oil tankers break down (ID73) and try to move out of shallow land (ID74) (cause 8). Bunker fuel floats on saltwater (ID75) and usually floats on freshwater (ID76) (description 14). It is black (ID77) and may be sticky for a time until it weathers (ID78), but even then it can persist in the environment for months (ID79) or even years if not removed (ID80) (effect 3). While this oil can be very long lasting (ID81) and has a greater tendency to adhere to surfaces (ID82) such as animal skins (ID83), fur (ID84), hair (ID85) or feathers (ID86), it is generally significantly less acutely toxic than lighter oils (ID87) (effect 4). All birds have what is called a preening gland (ID88) (elaboration 3- E). The preening gland secretes waxes (ID89) and fats that a bird spreads throughout its feathers (ID90) in order to make itself waterproof/insulated (ID91) (elaboration 4-E). Over the short term, if heavy oils get onto the feather (ID 92) and fur of birds (ID93) and animals, they find it harder to float in the water (ID94) and regulate their body temperatures (ID95) because oil destroys the insulating properties of fur (ID 96) and feathers (ID 97) (effect 5). This could lead to hypothermia (ID98) which could result in metabolic shock (ID99)(effect 6). That's why it is very important to keep animals (ID100) and birds away from contaminated areas until a spill is cleaned up (ID101) (remedy 4). To accomplish this, devices such as floating dummies are often used to scare them away from oily areas (ID102) (remedy 5). Over the long term, heavy oils could cause chronic health problems (ID103) such as tumors in some organisms (ID104) (effect 7). In cases where birds (ID105) and mammals become contaminated, rehabilitation centers are often set up to care for oiled animals (ID106) (remedy 6).

Non-sticky light oils (ID107) such as diesel fuel tend to be more toxic to organisms than heavy sticky oils (ID108) (effect 8). When spilled into water these oils spread quickly (ID109) and create a rainbow (ID110) or silvery sheen on the water (ID111) (effect 9). A sheen is a thin layer of oil less than 0.0002 inches (ID112) or 0.005 mm floating on the water's surface (ID113) (Effect 10). Spills usually occur from recreational (ID114) and fishing vessels that spill 500-5,000 gallons (ID115) (cause 9). Light oils are extremely toxic when they come into contact with marine life (ID116) (effect 11). For instance, mortality among birds due to ingesting it during preening (ID117) or by hypothermia from matted feathers is often reported (ID118)(effect 12).

Since light oil spills are nearly impossible to clean up (ID119) or contain (ID120), the best response to a spill is to leave it to disperse by natural means (ID121) (remedy 7). A combination of wind (ID122), sun (ID123), and wave action can rapidly scatter (ID124) and evaporate these oils (ID125) (remedy 8). While light oils can be dispersed naturally (ID126), dispersants are introduced to break up heavy oils (ID127) and speed natural biodegradation (ID128) (remedy 9). Dispersants are most effective when used within an hour (ID129) or two of the initial spill (ID130) (remedy 10). However, these materials may increase the harmful effects of oil on the insulation abilities of bird feathers (ID131)(effect 13). Dispersants could also cause oil particles to diffuse more deeply in the water column (ID132) where the oil may harm populations of animals in deeper waters (ID133) (remedy 11). In some cases, biological agents also are introduced to hasten biodegradation of oil components (ID134) (remedy 12). Biodegradation of oil is a natural process that slowly - sometimes over the course of several years - removes oil from marine environment (ID135) (remedy 13)

Other reported repercussions of oil spills include animal mortality (ID136) due to blockage of air passageways (ID137) (effect 14). Many baby animals (ID138) and birds starve to death (ID139) because their parents cannot detect the natural body scent of their offspring (ID140)(effect 15). Experts in bird (ID141) and mammal rehabilitation centers observe animals for signs of pneumonia from inhaling the toxic fumes of the oil (ID142)(remedy 14). They also check for stomach ailments that follow from ingesting the oil (ID143) (remedy 15).

Section 3

Land-based oil spills can originate from pipelines (ID144), storage facilities (ID145), fixed industrial facilities (ID146), and air transport facilities (ID147) (cause 10). These types of spills are usually localized (ID148) and thus their impact can be dealt with relatively easily (ID149) (effect 16). Environmental experts assert that most oil spills on land are preventable (ID150) (description 15). Following appropriate protocols (ID151) and maintenance procedures helps to ensure the safe operation of equipment (ID152), aiding in the goal of preventing spills from occurring (ID153) (description 16).

Spills could originate in underground (ID154) and above ground storage tanks that are used to store petroleum products (ID155) (cause 11). For example many homes built before 1957 in North America have underground heating oil tanks (ID156) and many businesses (ID157) and municipal highway departments also store gasoline (ID158), diesel fuel (ID59) and fuel oil in site-tanks (ID160) (cause12). Over the years, there have been reports that the contents of these underground tanks have leaked and spilled into the environment (ID161) (cause 13). Underground tanks corrode as they age (ID162) leading to a leak of oil (ID163) (cause 14). The common life expectancy of buried oil tanks is 10-15 years (ID164) (Elaboration 5-C). At about 20 years, the risk of leaks from buried steel oil tanks becomes significant (ID165) (cause 15). Leaks can occur earlier if a tank was damaged at installation (ID166) or was not properly piped (ID167)(cause 16).

The contents of underground tanks can contaminate the soil (ID168) (Effect 17). If a spill is severe it may take the land years to recover (ID169), during which time it is toxic to many plants (ID170) and animals (ID171) (effect 18).

Corrosion is not the only problem storage tanks can suffer leading to an oil leak (ID172) (cause 17). Problems with valves could also lead to a spill (ID173) (cause 18). First responders should identify the source of the oil (ID174) and where it is leaking to prevent further leakage (ID175) (remedy 16). Once the problem is identified, every attempt should be made to stop the release of oil (ID176), either by closing valves (ID177) or deflecting the oil into a containment area (ID178) (remedy 17).

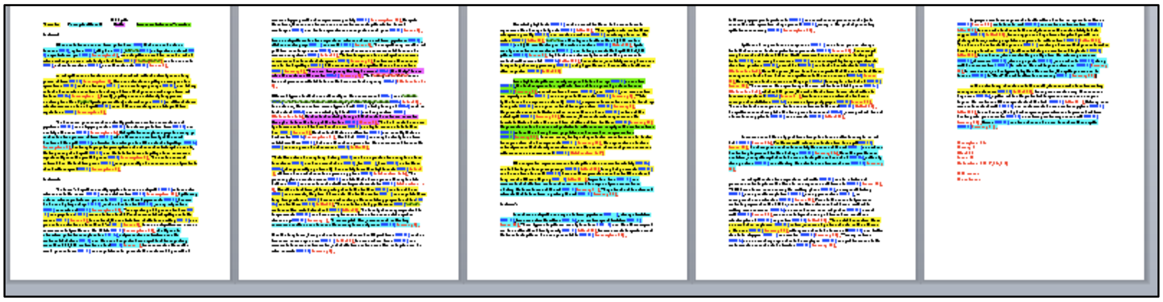
An oil spill could also originate in oil wells (ID179) due to a failure of pressure control systems that lead to underground oil blowouts (ID180) (cause 19). Well blowouts can occur during the drilling phase (ID181), during well testing (ID182), during well completion (ID183), during production (ID184), or during work over activities (ID185) (cause 20). Prior to the advent of pressure control equipment in the 1920s, uncontrolled releases of oil from a well while drilling were common (ID186) and were known as an oil gusher (ID187) or wild well (ID188) (cause 21) An accidental spark during a blowout can result in a catastrophic oil (ID189) or gas fire (ID190) (effect 19). Thus until it is certain there is no risk of an explosion (ID191) or fire, sources of ignition should not be allowed in the area (ID192) (remedy 18). All engines should be turned off (ID193) and traffic should be stopped (ID194) or diverted (ID195) (remedy 19). Warning notices (ID196) and no smoking signs should be displayed (ID197) and public access to the contaminated area should be limited (ID198) (remedy 20).

Improper waste management of oil facilities located on agricultural lands (ID199) (cause 22) could lead to soil (ID200) and water contamination (ID201) (effect 20) In case oil is leaked into the soil, it reduces the soil's ability to hold oxygen (ID202) (effect 21). Since oxygen is a key element in plant growth and photosynthesis, existing vegetation is prone to suffocation (ID203) (effect 22). The oil can also saturate the ground (ID204) creating a barrier that prevents water from being absorbed (ID205), further inhibiting nourishment to plants (ID206) (effect 23). In such case, it is essential to quickly clean the oil using vacuums (ID207), skimmers (ID208), cleaning agents (ID209), and sorbents (ID210) along with bioremediation (ID211) to return the soil to productive use (ID212) (remedy 21). In some cases, as it is typically lighter than water, the spillage area can be flooded (ID213) to "float" the oil to allow for recovery (ID214) (remedy 22).

As the oil is being absorbed into the earth, it causes risk of contaminating underground streams (ID215) (effect 24). Since groundwater may flow across a large area (ID216), spilled oil has the potential to spread across a wide region beyond the confines of the original site of the leak (ID217) (effect 25). Oil may even come into contact with (ID218) and contaminate human water supplies (ID219) (effect 26). In such cases, the highest response priorities are to prevent oil from leaching into groundwater (ID220) or entering waterways as run-off (ID221) (remedy 23). Berms (ID222) and trenches can be used to contain these spills (ID223) (remedy 24).

Appendix B. Achievement Tests and Rubrics

Color-coded screenshot to show how different parts of text are mapped to both transfer and prompted recall questions



Transfer questions

1. Which of these incidents is considered an oil spill.

   a. *An oil discharge was recorded in a report described that 4 ppm of oil was discharged over a period of two years from an oil refinery.*

   b. *Routine surveillance of shipping routes detected slight discharges from a British tanker handling cargo*

   c. *An oil vessel ruptured near California coast, spilling 100 tons in 24 hours.*

   d. *An off shore well reported discharges of 3 ppm over a period of one year.*

   e. *Discharges from a cargo vessel while taking in fuel.*

2. Choose the option that best describes the current situation regarding oil spills:

   a. *Massive social protest, incorrect public perceptions about the frequency of oil spills, and a lot of oil spills incidents*

   b. *Moderate social protest, right public perceptions about frequency of oil spills, and a lot of oil spills incidents*

   c. *Massive social protest, incorrect public perceptions about the frequency of oil spills, and few oil spills incidents*

d. *Moderate social protest, incorrect public perceptions about the volume of oil spillage and few oil spills incidents*

e. *Massive social protest, correct public perceptions about the volume of oil spillage and a lot of oil spills incidents*

3. You are in charge of 21 staff who survey tankers, pipelines and offshore drilling sites. Choose the best way to divide up your staff into teams.

    a. *7 in tankers, 7 pipelines and 7 offshore drilling*

    b. *13 off shore drilling, 4 tankers and 4 pipelines*

    c. *9 tankers, 9 pipelines and 3 offshore drilling*

    d. *2 tankers, 2 pipelines and 17 offshore drilling*

    e. *15 tankers, 2 pipelines and 4 offshore drilling*

4. The text described four ways to deal with marine oil spills. Name each remedy and for each, provide an example not mentioned in the text where it is the best way to respond to an oil spill.

5. A cargo ship is sailing through a group of islands shallow waters. The captain reported mechanical issues but help was not provided. Although no oil spill was reported, an ecology activist group in the area campaigned against using this route to ship oil. Is the ecology activist right to be concerned? Explain.

6. There is a spill of 50 barrels of bunker fuel in a harbor in Town A. In Town B, 50 barrels of diesel fuel spilled in a marina. After 6 months, which spill will still be visually present?

7. Jet skis and speedboats at marine resorts discharge diesel fuel into the ocean but they are not prosecuted. Ecological activists claim it is too hard to find proof of wrongdoing in these cases. Why is this?

8. The media fiercely attacked the U.S. Environmental Protection Agency for not doing enough about a diesel fuel spill from an ocean-going vessel in the Gulf of Mexico. Explain why there was a lack of action.

9. Dispersants were used in response to a heavy oil spill in the Red Sea, home to many species of corals. Is this a good response? Explain your answer.

10. One oil tanker is carrying 200,000 barrels of gasoline. A second tanker is carrying 500,000 barrels of bunker oil. These two tankers collided in a habitat where sea otters live. Given what you know about harmful effects of oil spills on birds, what are potential negative consequences of this oil spill on sea otters?

11. A town founded in Quebec in 1950 reported significant reductions in farm incomes and poor plant growth over the past 5 years. An analysis of the soil showed oil residue. What could be the cause?

12. A new development of homes is next to a natural gas well. Which is the most important item on a safety inspector's list:

    a. Exterior lighting

    b. Barbeques with electric starters

    c. Glass windows

    d. Electric cable insulation

    e. Exit doors

13. 5000 barrels of oil leaked into the soil near a forest populated by many animals. Although scientists reported normal air quality in the contaminated area, there is a particularly high rate of mortality among animals especially elk, brown rabbits and grizzly bears. What could be the possible reasons for this profile of effects of the spill?

Prompted recall questions

1. What is oil?

2. Why might oil spilled from a ruptured pipeline not be noticed quickly?

3. What substances may be spilled in an oil spill?

4. List sources of accidental marine oil spills.

5. Why were marine oil spills a serious problem in the decade 1960-1970?

6. What are containment booms?

7. Why are floating dummies used in some oil spills?

8. What is sheen?

9. Following an oil spill, what is the most important thing experts need to check when caring for animals in rehabilitation centers?

10. List sources of land based oil spills.

11. What should be done when a leak from an underground tank is detected?

12. What could happen if waste management of oil facilities located on agricultural lands is not done well?

13. Describe a situation in which each of these responses to an oil spill is appropriate. How does the response reduce negative consequences of the oil spill?

    a. Skimmers

    b. Berms

    c. Flooding with water

Transfer Questions Rubric

1. C (1pt.) (ID 8, ID 12, ID 13, ID 14, ID 15, ID 16, ID 17, ID 18, ID 19)

2. C (1pt.) (ID 27, ID 28, ID 29, ID 30)

3. C (1pt.) (ID 40, ID 41, ID 42, ID 43, ID 44)

4. Biological agents (1pt.) example based on: “to hasten biodegradation of oil components when natural biodegradation is slow” (1pt.) (ID 134, ID 135), dispersants (1pt.) example based on: “to break heavy oils is effective when used after an hour or two of the initial spill OR but should not used in context with birds because it increase harmful effects of oil on insulation abilities of bird feathers” (1pt.) (ID 127, ID 129, ID 130, ID 131, ID 32, ID 133), containment booms (1pt.) example based on: “to limit spill if crew reaches spill within an hour or two to box in oil on surface (1pt.) (ID 58, ID 59, ID 60, ID 61, ID 62), disperse naturally (1pt.) example based on: with light oils when it is impossible to clean up or contain (1pt.) (ID 119, ID 120, ID 121, ID 122, ID 123, ID 124, ID 125).

5. Yes, they need to be concerned (1 pt.). When tanker breaks down and tries to move out of shallow water, fuel could be released through a hole caused by abrasion. (2 pt.) (ID 72, ID 73, ID 74).

6. Bunker fuel in town A (1 pt.) (ID 81 =ID 79, ID 80), ID 82, ID 83, ID 84, ID 85, ID 86,ID 87)

7. When diesel fuel spills into water, it spreads quickly thus cannot be detected (1 pt.) (ID 109).

8. Light oils are impossible to clean (1 pt.) or contain (1 pt.), best response is to leave to disperse naturally (1 pt.) (ID 119, ID 120, ID 121).

9. Not a good response (1 pt.).

   They could cause oil particles to diffuse more deeply in the water column (1 pt.) where oil could harm animals in deeper waters (1 pt.) (ID 132, ID 133).

10. Heavy oil: gets into their fur/ skin making hard to float in water (1 pt.), regulate their body temperature because it destroyers the insulating properties of fur and lead to hypothermia which could lead to metabolic shock (1 pt.) (ID 92, ID 93, ID 94, ID 95, ID 96, ID 97, (ID 98, ID 99 = ID 118, ID 117)

    Light oil: animal mortality due to blockage of air passageways, they suffer from pneumonia from inhaling toxic fumes (1 pt.), ingesting the oil causes mortality (1 pt.) Both oil types: baby animals starve to death because their parents cannot detect the natural body scent of their offspring (1 pt.). (ID 136, ID 137, ID 138, ID 139, ID 140)

11. Many homes build before 1957 in North America had underground tanks for heating where spills could originate (1 pt.). These tanks corrode as they age leading to a leak of oil about 20 years the risk of leaks becomes significant (1 pt.). Leaks could occur earlier if the tank is damaged at installation (1 pt.). Problems with valves could also lead to leaks (1 pt.). (ID 156, ID 157, ID 158, ID 159, ID 160,ID 161, ID 162, ID 163, ID 165, ID 166, ID 167, ID 173)

12. B (1 pt.).(ID 191, ID 192)

13. Soil contamination: when oil is leaked into the soil it reduces soil's ability to hold oxygen (1 pt.), leading to suffocation of existing vegetation since oxygen is important for photosynthesis (1 pt.) (ID 202, ID 203). Oil can act like a barrier stopping water from reaching the plant, thus it dies (1 pt.), when there are no plants animals that are plant eaters (rabbits and elks) won't find anything to eat (1 pt.) (ID 204, ID 205, ID 206). When elk and rabbits die, grizzly bears have nothing to eat (1 pt.) Water contamination: when oil is absorbed into the soil it could contamination underground streams so animals could die from contaminated water (1 pt.) (ID 215).

Prompted Recall Rubric

1. Oil is a substance derived from petroleum (1 pt.) (ID 1) that does not include benzene (1 pt.) (ID 2), xylene (1 pt.) (ID 3), ethylene (1 pt.) (ID 4), toluene (1 pt.) (ID 5) or liquefied natural (1 pt.) (ID 6) or petroleum gas (1 pt.) (ID 7).

2. When there is uncertainty about the location of the spill (1 pt.) and limited knowledge about the extent of the spillage (1 pt.).

3. Refined petroleum products (1 pt.) (ID 34), their by products (1 pt.) (ID 35), heavier fuels by large ships (1 pt.) (ID36) such as bunker fuel (1 pt.) (ID 37), oil refuse (1 pt.) (ID 38) waste oil (1 pt.) (ID 39).

4. Pipelines (1 pt.) (ID 53), offshore drilling rigs (1 pt.) (ID 54) and wells (1 pt.) (ID 55). Bunker fuel could be released through a hole in the body of a tanker (1 pt.) (ID 73). Recreational (1 pt.) (ID 114) and fishing vessels (1 pt.) (ID115).

5. Because of intensified petroleum exploration (1 pt.), oil production activities (1 pt.), and the use of supertankers capable of transporting more than 500,000 metric toms of oil (1 pt.).

6. They are temporary floating barriers (1 pt.) (ID 61) that help box in oil on the surface of the water (1 pt.) (ID 62).

7. To keep (scare) animals and birds away from contaminated areas (oily areas) (1 pt.).

8. A thin layer of oil floating on the water's surface (1 pt.) (ID 112) less than 0.0002 inches or 0.005 mm (1 pt.) (ID 113).

9. Experts observe animals for signs of pneumonia from inhaling the toxic fumes of the oil (1 pt.) (ID 142), they also check for stomach ailments that follow from ingesting the oil (1 pt.) (ID 143).

10. Pipelines (1 pt.) (ID 144) storage facilities (1 pt.) (ID 145), fixed industrial facilities (1 pt.) (ID 146), air transport facilities (1 pt.) (ID 147).

11. Identify the source of the oil and where it is leaking (1 pt.) (ID 174 OR ID 175)

    Stop the release of oil by closing valves (1 pt.) (ID 176 Or ID 177)

    Deflecting oil into containment area (1 pt.) (ID 178)

12. Lead to soil (1 pt.) (ID 200) and water contamination (1 pt.) (ID 201)

13. a. Skimmers: clean the oil (1 pt.) (ID 207), b. berms: contain the spill) (1 pt.) (ID 223), c. flooding with water: to float the oil and allow for recovery (1pt.) (ID 214).

Appendix C. FluidSurvey Workflow

1. Demographic questionnaire: Sex, Major and age   Click ==➔*Next*

2. Prior knowledge: ***"****List in bullet form 7 to 8 important ideas they know about oil spills, and causes of, effects and remedies for oil spills."* Click ==➔*Next*

3. MSLQ: (General) Selected sections of MSLQ adapted to fit this study: Metacognitive self-regulation, critical thinking, rehearsal, elaboration (see Appendix). Click ==➔*Next*

4. Reading text: Study text (each group will have specific instructions to follow while reading and studying). Click ==➔*Next*

5. MSLQ: (Specific) Selected sections of MSLQ adapted to fit this study: Metacognitive self regulation, critical thinking, rehearsal, elaboration (see Appendix) Click ==➔*Next*

6. Inferential questions: 13 questions. (See Appendix for mapping of questions to text parts) Click ==➔*Next*

7. Prompted recall questions: 13 questions. (See Appendix for mapping of questions to text parts) Click ==➔*Next*

8. Only for "*Free marking*" group: Show the text with highlights they made and ask them to write why they highlighted these parts? What criteria did they use to highlight?   Click=➔ *Submit*

Appendix D. Correlations of MSLQ, Performance and Marking

Groups given Criteria

Correlations of prompted recall, transfer, and Ratings on MSLQ General and MSLQ Specific

	Total MSLQ-G.	Total MSLQ-S.	Total P. Recall	Total Transfer
Total MSLQ-G.	1	0.595**	-0.14	-0.143
Total MSLQ-S.		1	-0.213	-0.225
Total P. Recall			1	0.609**
Total Transfer				1

Note: ** p<0.01; *p < 0.05

Marking Groups

Correlations of prompted recall, transfer, and Ratings on MSLQ General and MSLQ Specific

	Total MSLQ-G.	Total MSLQ-S.	Total P. Recall	Total Transfer
Total MSLQ-G.	1	0.330**	0.077	-0.089
Total MSLQ-S.		1	0.119	0.057
Total P. Recall			1	0.604**
Total Transfer				1

Note: ** p<0.01; *p < 0.05

Correlations of Marking Descriptions, Marking Causes, Marking Effects, Marking Remedies and Ratings on MSLQ General and MSLQ Specific

	Total MSLQ-G.	*Total MSLQ-S.*	Marking descript.	Marking causes	Marking effects	Marking remedies
Total MSLQ-G.	1	0.330*	0.183	-0.031	0.098	0.271
Total MSLQ-S.		1	0.149	0.04	0.074	0.106
Marking descript.			1	0.897**	0.894**	0.281
Marking causes				1	0.920**	0.144
Marking effects					1	0.24
Marking remedies						1

Note: ** p<0.01; *p < 0.05