7-1-2011

Generating Keywords Improves Metacomprehension and Self-Regulation in Elementary and Middle School Children

Anique B.H. de Bruin
Maastricht University

Keith Thiede
Boise State University

Gino Camp
Erasmus University Rotterdam

Joshua Redford
Boise State University

The final, definitive version of this document can be found online at *Journal of Experimental Child Psychology*, doi: 10.1016/j.jecp.2011.02.005

Generating Keywords Improves Metacomprehension and Self-Regulation in Elementary and Middle School Children

Anique B.H. de Bruin
Maastricht University

Keith W. Thiede
Boise State University

Gino Camp
Erasmus University Rotterdam

Joshua Redford
Boise State University

Abstract

Metacomprehension accuracy is typically quite poor; however, recently interventions have been developed to improve accuracy. In two experiments, we evaluated whether generating delayed keywords prior to judging comprehension improved metacomprehension accuracy for children. For sixth and seventh graders, metacomprehension accuracy was greater for the delayed-keyword condition than for a control group. By contrast, for fourth graders, accuracy did not differ across conditions. Improved metacomprehension accuracy led to improved regulation of study.

Keywords: metacomprehension accuracy, self-regulation, text comprehension, development

Models of self-regulated learning suggest that learning involves both metacognitive monitoring and control, or regulation of study (e.g., Ariel, Dunlosky, & Bailey, 2009; Nelson & Narens, 1990; Thiede & Dunlosky, 1999; Winne & Hadwin, 1998). To illustrate, consider a student preparing for an upcoming examination: The student has a goal of mastering the material. As she reads her science textbook, she monitors how well she has understood the various sections of the chapters. This monitoring then informs her regulation of study, where she chooses to reread those sections in which mastery has not yet occurred. Within this context, more accurate monitoring leads to more effective regulation of study, and this in turn leads to better comprehension (Thiede, Anderson, & Therriault, 2003). As Nelson and Narens (1990) describe it, cognition (attention, learning, problem solving) takes place at an object level, whereas people also have a meta-level containing a model of a student's understanding of the learned material. Monitoring takes place when information from the object level flows to the meta-level and informs it about what is happening at the object level, for instance when a student realizes her understanding of a chapter falls short. Control (often called regulation) is exerted when the meta-level acts to influence the object level, for example by deciding to reread a section of a chapter. The interdependence between monitoring and regulation becomes clear in this example: Monitoring will only influence learning when it leads to regulation of study behavior.

Interest in metacognition and self-regulated learning has been on the rise since the 1980s (for recent books providing overviews see Dunlosky & Metcalfe, 2009; Hacker, Dunlosky, & Graesser, 2009). Moreover, interest in comprehension monitoring (called metacomprehension) and self-regulation during reading has also increased in recent years (for reviews of this literature see Dunlosky & Lipko, 2007; Maki, 1998; Thiede, Griffin, Wiley, & Redford, 2009). In the broader area of metacognitive monitoring, compared to research on college students, less research has been done with children. Most important, this is the first study to examine how well children monitor their comprehension during reading or whether children use their monitoring to regulate subsequent rereading. To provide a theoretical framing for the present research, we will provide a brief review of the literature on children's metacognitive monitoring and regulation of study. However, first we will provide an overview of the paradigm used to measure metacomprehension accuracy and describe how it has been operationally defined in the literature.

Glenberg and Epstein (1985) developed the paradigm that is commonly used to measure metacomprehension accuracy. In this paradigm, participants read a set of texts, judge how well they think they have understood each text, and then take a test for each text. Metacomprehension accuracy is defined as the intra-individual correlation between metacomprehension judgments and actual test performance computed across the texts (Nelson, 1984). It is important to note that metacomprehension accuracy (also called relative accuracy or resolution, Weaver, 1990) describes a person's ability to discriminate between texts that were better learned versus those that were less learned. Although a number of studies have examined children's monitoring accuracy in less complex tasks, the present study is the first to look at metacomprehension accuracy of children. We will review the studies on monitoring accuracy in children next.

Monitoring Accuracy

The majority of research investigating the relative accuracy of metacognitive monitoring has involved college students. However, recently researchers have examined whether the most robust effects in monitoring accuracy extend to children.

Schneider, Visé, Lockl, and Nelson (2000) were the first to apply the classic metamemory paradigm to children, to evaluate whether the delayed judgment of learning (JOL) effect extended to children. The delayed-JOL effect (Nelson & Dunlosky, 1991) refers to the finding that monitoring accuracy improves dramatically when JOLs are made after a delay rather than immediately after studying a word pair (for additional theoretical discussion of the mechanisms producing the effect see Dunlosky & Nelson, 1997; Spellman & Bjork, 1992). Schneider et al. (2000) reported results that mirrored those previously found in adults: children of 6, 8, and 10 years old more accurately monitored memory performance when JOLs were made after a delay. Son (2005, 2010) also showed that children had low but better than chance levels of accuracy for immediate JOLs. Koriat and Shitzer-Reichert (2002) showed that, just as adults, second and fourth graders take into account item difficulty when giving JOLs: Children provided higher JOLs for easy word pairs than for difficult word pairs. Second graders, however, overestimated performance on the difficult items. As in adults, JOL accuracy improved over learning trials (note that with younger children (ages 4 – 5), practice did not improve monitoring accuracy, Lipko, Dunlosky, & Merriman, 2009).

These studies show that as of elementary school children are able to monitor their own learning to a certain extent. However, all of the aforementioned studies have made use of less complex learning materials that may not be comparable to what is studied more commonly in school. To date, no one has examined how accurately children monitor comprehension during reading.

One reason to suspect that children may struggle to monitor their own learning in a more complex task, like reading, is that learning these materials may demand more cognitive resources than more simple tasks, which may leave fewer resources for monitoring learning (Rawson, Dunlosky, & Thiede, 2000). This could be particularly important with children, as Roebers, von der Linden, and Howie (2007) showed that cognitive resources play an important role in children's monitoring.

Regulation of Study

Equally important to learning as accurate metacognitive monitoring is effective metacognitive control or regulation of study. Effective regulation would involve allocating more study time to difficult items than to easy items (see Nelson, Dunlosky, Graf, & Narens, 1994; Thiede, 1999). Thus, more effective regulation of study is indicated by a stronger negative correlation between JOLs and allocation of study time.

Research in regulation of study is inconsistent regarding whether children are capable of allocating study time efficiently over study materials. In a study by Masur, McIntyre, and Flavell (1973) seven-year-olds, nine-year-olds, and college students studied a list of pictures and were then required to select half of the list for restudy. Whereas the nine-year-olds and college students tended to correctly select the items they did not recall on the first trial, the seven-year-olds chose items randomly for restudy. However, in this study, students were forced to restudy items, which might have led to unnatural restudy behavior. Dufresne and Kobasigawa (1989) examined spontaneous study time allocation in 6- to 12-year olds. Older children (10- and 12-year old) distributed study time effectively, studying difficult items longer than easy items, whereas younger children (6- and 8-year old) distributed study time evenly across difficult and easy items. Post-experimental questions revealed that even the youngest group differentiated between the difficult and easy items, which led the authors to conclude that memory monitoring shows little developmental trends, but self-regulation of learning does. However, later research showed that the ability to self-regulate study depends largely on the difficulty of the task. Kobasigawa and Metcalf-Haggart (1993) had first and third grade children study familiar and unfamiliar objects until they were able to name all of the objects correctly. Given the salient difference in item difficulty in this task, even the first grade students dedicated more time to study of the unfamiliar items.

None of the abovementioned studies, however, examined how monitoring affects self-regulation of study in children. As emphasized by Nelson and Narens (1990), monitoring and regulation of study are interdependent; Without accurate monitoring, regulation of study will fail. In an attempt to study the interdependent relation between monitoring and regulation in children, Lockl and Schneider (2003) had seven- and nine-year-old children study easy and difficult paired associates. Students then provided JOLs and studied the word pairs once more at their own pace. The results showed that both age groups provided higher JOLs for the easy than for the difficult word pairs and devoted more study time to difficult items. The gamma between JOLs and study time of the older children, however, was higher (gamma = -.40) than that of the younger children (gamma = -.22) indicating that the ability to allocate study time based on individual JOLs increases with age. Also, children who self-regulated study effectively had higher recall performance on the test than those who were poor self-regulators.

Son (2010) examined the relation between metacognitive monitoring and choice of study strategy (whether to mass or distribute study of word pairs). Children (grades 3 – 5) used their metacognitive monitoring to guide study strategy, but were less effective than were adults. Son (2005) found that, in contrast to adults, first graders generally favored massed study and selected this strategy indiscriminately across items. These findings suggest a developmental trend in strategy selection.

Roebers, Schmid, and Roderer (2009) had third and fifth graders rate the correctness of their answers on a test about a previously learned science topic and provided them the opportunity to regulate their test taking behavior by letting them withdraw answers they were unsure about. Results showed that both age groups were able to accurately monitor their performance on the test as indicated by the high gammas (.63 to .75), and that the students showed emerging control skills. They more often withdrew incorrect than correct answers, and their JOLs were lower for the withdrawn answers. Although both age groups withdrew incorrect answers to the same extent, the third graders more often withdrew correct answers than the fifth graders, possibly indicating a stricter response criterion. In a related study by Krebs and Roebers (2010), they examine how retrieval processes influence monitoring and control behavior. They found that children more often withdrew incorrect answers from unanswerable items than from answerable items, indicating that memory retrieval does play an important role in monitoring and controlling test taking behavior. This effect was more pronounced for the fifth than for the third graders, which emphasizes that the use of retrieval processes in monitoring and control is dependent on age.

In sum, the literature suggests that skills involved in metacognitive monitoring are fairly well developed by second grade. By contrast, being able to use this monitoring to guide subsequent study (regulation of study) seems to develop later.

Linking Metacomprehension Accuracy and Regulation of Study

Models of self-regulated learning suggest that accurate monitoring of learning is critical to effective regulation of study, and thus, to learning. However, the vast majority of evidence linking monitoring to regulation of study is correlational; as is the evidence linking monitoring and regulation to learning. One exception is a study conducted by Thiede et al. (2003). As we used their procedure in the present experiments, we will describe the method here. Thiede et al. (2003) altered monitoring accuracy by manipulating keyword generation across three groups. Because they used texts as study material, this is generally referred to as metacomprehension accuracy rather than monitoring accuracy. In particular, a no-keyword group completed the standard metacomprehension procedure. They read a set of texts, judged their comprehension of each text (after reading all the texts), and then completed a test for each text. A delayed-keyword group read all the texts, after which they generated a list of five keywords that captured the essence of each text. They then judged their comprehension of each text, and finally completed a test for each text. An immediate-keyword group read a text and immediately generated a list of five keywords for the text. They then proceeded to the next text. After reading and generating keywords for each text, they then judged comprehension for each text and finally completed the test for each text. Metacomprehension accuracy was calculated as a gamma correlation between comprehension judgments and test performance. Thiede et al. (2003) showed that metacomprehension accuracy was dramatically greater for the delayed-keyword group than for the other groups, which did not differ in accuracy (for more theoretical discussion of the keyword effect and additional evidence of the delayed-keyword effect see Thiede, Dunlosky, Griffin, & Wiley, 2005).

The boost in metacomprehension accuracy led to more effective regulation of study. That is, participants (college students) in the delayed-keyword group allocated more study time to texts that were less learned than to texts that were better learned to a greater extent than did participants in the other groups. Moreover, overall comprehension test performance was significantly greater for the delayed-keyword groups than for the other groups. Thiede et al. (2005) propose that generating keywords at a delay after text study activates the situation model, which is subsequently used as a cue for comprehension judgments. The situation model is a mental representation of the text containing the connections among the ideas of a text and a person's prior knowledge (see Kintsch, 1998, for more on this and other representations presented in the construction-integration model of text comprehension). As the situation model is a strong predictor of comprehension test performance (McNamara, Kintsch, Songer, & Kintsch, 1996; Wiley, Griffin, & Thiede, 2005), metacomprehension accuracy improved in the delayed keyword group (for an empirical test of the situation model hypothesis compared to an accessibility explanation, see Anderson & Thiede, 2008). As described in models of self-regulated learning (Dunlosky & Hertzog, 1997; Metcalfe, 2002), the more accurate metacomprehension of the delayed keyword group led to better regulated restudy, and ultimately, to better performance on the test.

Overview of Experiments

The goal of the present experiments was to examine the development of metacomprehension accuracy and regulation of learning in children. Given the positive effect found in previous studies (Thiede et al., 2003, 2005), we used the no-keyword and delayed-keyword conditions in the present experiments. As in the literature, metacomprehension accuracy was operationally defined as the gamma correlation between an individual's comprehension judgments and his or her test performance, computed across the set of texts (for a rationale for using gamma, see Nelson, 1984). If the delayed-keyword effect extends to younger children, we expect to see metacomprehension accuracy greater for the delayed-keyword group than for the no-keyword group.

After completing the final test in the present experiments, participants were asked to select texts they would like to reread. As recommended by Nelson (1984) and done in previous metacomprehension research (e.g., Thiede et al, 2003), self-regulation was operationally defined as the gamma correlation between an individual's metacomprehension judgments and his or her selection of texts for reread (where a 1 indicates a text was selected and 0 indicates a text was not selected for rereading), computed across the set of texts. More effective regulation of study is indicated by a stronger negative correlation. If children are able to regulate their study behavior, we expect to see more effective regulation of study for the delayed-keyword group than for the no-keyword group—because the superior metacomprehension accuracy should lead to better regulation of study, as in Thiede et al. (2003).

Experiment 1

Method

Participants and Design. Ninety-four seventh grade students (ages 12 – 13) participated in the experiment; they were from the Northwestern part of the United States. Of the 94 participants 51 (54%) were female and 43 (46%) were male. Participants completed the experimental session in the classroom setting, with their regular teacher supervising along with a researcher. All participants were treated in a manner consistent with the ethical standards of the American Psychological Association. Participants were randomly assigned to one of two groups: The delayed-keyword group generated keywords prior to judging their comprehension of texts; the no-keyword group served as a control group.

Materials. The texts were five expository science texts that described a complex causal relation. Texts covered topics from the seventh grade life science curriculum (e.g., photosynthesis, cellular respiration), which had not yet been covered in class. A sample text and test questions can be found in Appendix A. As suggested by Wiley et al. (2005), the texts were developed so that important elements of the situation model were not explicit in the surface form of the text (i.e., the causal connections among ideas in the texts were not stated and needed to be generated by the reader). Each text was approximately 350 words long—two teachers verified that the reading level was appropriate for students. For each text, we constructed a five item test that required participants draw inferences about the ideas presented in the text and was designed to assess the situation model of the text.

Procedure. All the participants were instructed that they would be reading a series of texts, judging their comprehension of each text, and then taking a test for each text. They were also instructed that they would be selecting texts for restudy, but that due to time constraints they would not actually read the selected texts. Participants were also instructed that they might be asked to write a list of keywords that captured the essence of a text. These instructions included an example of keywords (i.e., for a text on the Titanic one might write: iceberg, shipwreck, tragedy, etc.). Due to the straightforward nature of generating keywords, no formal training on how to generate keywords was provided. Participants were given an opportunity to ask questions about the procedure prior to starting experimental procedure. Through the experiment, participants were given as much time as needed to read texts, make judgments, complete tests, and select texts for restudy.

For the critical trials, the order of text presentation was randomized for each participant using a Latin Square design. For each participant, the order of presentation was maintained throughout the experimental procedure (e.g., reading, judging comprehension, and testing). Participants in the no-keyword group first read the five texts. After reading, they rated their comprehension for each text. The comprehension rating was prompted with the query, "Please circle how many of the five test questions you think you will answer correctly on the text entitled: TITLE." Participants selected the number of items they believed they would correctly answer: 0 to 5. After rating their comprehension of the last text, they answered five questions for each text. Participants in the delayed-keyword group read each of the five texts. They were then shown the title of a text and instructed to write five keywords that captured the essence of that text. Once they finished writing keywords for the last text, they judged their comprehension (using the prompt described above). After rating their comprehension of the last text, they answered the test questions for each text. After completing the last test, self-regulation was prompted in both groups by presenting the titles of each text and stated above it: "Please circle the texts you would like to restudy". As the goal of this study was to examine the development of metacomprehension and regulation of study in children, participants were not asked to restudy the selected texts.

Results and Discussion

Metacognitive judgments and test performance. The primary focus of this investigation is metacomprehension accuracy; however, as metacomprehension accuracy is the relationship between metacognitive judgments and test performance, we first report data on these variables. For each participant, we computed the median metacognitive judgment and test performance across the five critical texts. We used the median because it is the recommended measure of central tendency for small sets of scores where extreme scores may have an undue influence on the mean (Gravetter & Wallnau, 1999). The mean of the medians was computed across participants in each group (see Table 1). The mean magnitude of metacomprehension judgments did not differ across groups, $t(92) < 1.00$, $p = .53$. Test performance did not differ across groups, $t(92) < 1.00$, $p = .85$.

Insert Table 1 about here

Metacomprehension Accuracy. As recommended by Nelson (1984), metacomprehension accuracy was operationalized as the intra-individual gamma correlation[1] between a person's metacomprehension judgments and test performance across the five critical texts. A stronger positive correlation indicates greater accuracy. As seen in Figure 1, metacomprehension accuracy for participants in the no-keyword group (*M* = -.01, *SEM* = .09) was not significantly different from zero, *t*(47) < 1.00, *p* = .91, which indicates there was not a relationship between predicted and actual performance. By contrast, metacomprehension accuracy for participants in the delayed-keyword group (*M* = .27, *SEM* = .09) was significantly greater than zero, *t*(45) = 3.10, *p* = .003, indicating a significant relation between predicted and actual performance. Metacomprehension accuracy was greater for the delayed-keyword group than for the no-keyword group, *t*(92) = 2.21, *p* = .03.

Insert Figure 1 about here

Regulation of Study. Regulation of study was operationalized as the intra-individual gamma correlation between a person's metacomprehension judgment and selection of a text for restudy (where 1 indicated a text was selected for restudy and 0 indicated that the text was not). Fourteen students had indeterminate gamma correlations due to invariance in item selection. A stronger negative correlation indicates more effective regulation of study, as this suggests the person is allocating additional study time to texts that were perceived as less learned (cf. complete-compensation hypothesis, Nelson & Leonesio, 1988). As seen in Figure 2, the improved regulation of study was more effective for the delayed-keyword group than for the no-keyword group, *t*(78) = 2.01, *p* = .05.

Insert Figure 2 about here

These results extend the delayed-keyword effect to seventh grade students, and show that the improved metacomprehension accuracy was associated with more effective regulation of study. In Experiment 2, we examined whether younger children (fourth and sixth graders), who have only just started education in text comprehension, are also able to benefit from the keyword instruction to improve metacomprehension. Testing a younger group with limited text comprehension experience allows for an analysis of the development of metacomprehension skills in children.

Experiment 2

Method

Participants and Design. Eigthy-five fourth grade students (ages 9 – 10) and 66 sixth grade students (ages 11 – 12) participated in the experiment; they were from the Southwestern part of the Netherlands. Of the 151 participants 72 (48%) were female and 79 (52%) were male. Participants completed the experimental session in a classroom setting, with their regular teacher supervising along with a researcher. All participants were treated in a manner consistent with the ethical standards of the American Psychological Association. Participants were randomly assigned to one of two groups: A delayed-keyword group or a no-keyword group.

Materials. Participants studied six expository Dutch texts, which were adapted from elementary school books for length and content. The texts were on average 306 words long (range between 293 and 326 words), and covered the topics 'animals' (bears, elephants, and monkeys) and 'countries or regions' (Mexico, Egypt, and South-East Asia). These topics were chosen so as to relate to students' interests as much as possible. Therefore, different text topics were chosen than in Experiment 1. The texts were similar in structure to the texts in Experiment 1: A topic (Cellular respiration, elephants, etc.) was introduced, after which conceptual information and concepts were explained relating to the topic. For reasons of comparability, the same texts were used for fourth and sixth graders. Teachers verified that the reading level was appropriate for students in both grades. That is, teachers ensured that both grades should be well able to comprehend the information in the texts. As in Experiment 1, five inference test questions were designed per text. These questions did not refer to literal information in the text, but required connecting at least two ideas from the text. A sample text and test questions can be found in Appendix B.

Procedure. The procedure used in this experiment was similar to that in Experiment 1, with one exception. As reading 5 consecutive texts would have been too much for these younger students, we conducted the second experiment in two sessions. To equate the number of texts per session, one more text was studied compared to Experiment 1- with three texts per session.

All the participants were instructed that they would be reading a series of texts, judging their comprehension of each text, and then taking a test for each text. The experimenter verified that participants understood how to make metacomprehension judgments and how to select texts for restudy. Participants of both grades were randomly assigned to one of two groups (no-keyword or delayed-keyword). The no-keyword and delayed-keyword groups were similar to those in Experiment 1.

Each experimental session covered three texts of the same domain. Order of the domain (animals versus countries and regions) and texts was counterbalanced between participants. In the second session, the experimenter repeated the instructions, with referral to the previous session. There was no effect of session or order, so data from the two sessions were combined for analysis.

Results and Discussion

Metacognitive judgments and test performance. For each participant, we computed the median metacognitive judgment and test performance across the six critical texts. The mean of the medians was computed across participants in each group (see Table 1). Analysis of variance on median metacomprehension judgments showed no main effect of group or grade, $F(1, 147) = 2.81$, $MSE = .67$, $p = .096$, and $F < 1$, respectively. The interaction effect was significant, $F(1, 147) = 4.06$, $MSE = .67$, $p = .046$. Post hoc tests revealed that the fourth graders provided higher metacomprehension judgments in the no-keyword group than in the delayed-keyword group, $p < .01$. Metacomprehension judgments of the sixth graders did not differ between the no-keyword and delayed-keyword groups, ANIQUE, I would probably report the F rather than the p-value, just to be consistent, $F < 1$.$p = .82$. Test performance did not differ between groups, $F(1, 147) = 1.91$, $MSE = .99$, $p = .17$, but did between grades, $F(1, 147) = 9.47$, $MSE = .99$, $p = .002$. The fourth graders performed worse on the test than the sixth graders.

Metacomprehension accuracy. As in Experiment 1, metacomprehension accuracy was calculated as the intra-individual gamma correlation between a person's metacomprehension judgments and test performance across the six critical texts. As shown in Figure 3, metacomprehension accuracy for fourth and sixth graders in the no-keyword group was not significantly different from zero, $t(38) = 1.78$, $p = .08$ for fourth graders ($M = .18$, $SEM = .10$), and $t(32) < 1.00$, $p = .97$ for sixth graders ($M = .00$, $SEM = .11$). This shows that there was no relationship between predicted and actual performance in the no-keyword groups. However, with regard to the delayed-keyword groups, metacomprehension accuracy was significantly greater than zero for the sixth graders, $t(32) = 5.16$, $p < .001$ ($M = .42$, $SEM = .11$). For fourth graders, this was not significant, $t(45) = 1.04$, $p = .30$ ($M = .11$, $SEM = .09$). Comparing the no-keyword and delayed-keyword group between grades with an analysis of variance revealed no effect of grade, $F < 1$, or keyword, $F(1, 147) = 2.86$, $MSE = .39$, $p = .09$. However, a significant interaction between grade and keyword group was found, $F(1, 147) = 5.48$, $MSE = .39$, $p = .02$. Figure 3 shows that for fourth graders, there was no difference in metacomprehension accuracy between the no-keyword and the delayed-keyword group, $t(83) = .46$, $p = .64$, but for the sixth graders there was. They had higher metacomprehension accuracy in the delayed-keyword group than in the no-keyword group, $t(64) = 2.94$, $p = .01$. These results show that the positive effect of generating keywords extends to sixth but not fourth graders. Moreover, both sixth and fourth graders were unable to monitor text comprehension when not generating keywords, as indicated by the non-significant gamma's in the no-keyword group.

Insert Figure 3 about here

Regulation of Study. As in Experiment 1, regulation of study was operationalized as the intra-individual gamma correlation between a person's metacomprehension judgment and selection of a text for restudy (where 1 indicated a text was selected for restudy and 0 indicated that the text was not). Twelve children did not select texts for restudy. For these children, no regulation of study gammas could be computed. Analysis of variance with grade and keyword group as between-subjects factors showed a main effect of keyword group, *F*(1, 135) = 9.60, *MSE* = .48, *p* = .002. As seen in Figure 4, regulation of study was significantly better when generating keywords compared to not generating keywords. No effect of grade, *F*(1, 135) = 2.84, *MSE* = .48, *p* = .10, and no interaction effect was found, *F* < 1. Mean regulation of study gamma for the no keyword group was -.10 (*SEM* = .12) for fourth graders, and -.23 (*SEM* = .12) for sixth graders. Mean regulation of study gamma for the delayed keyword group was -.40 (*SEM* = .11) for fourth graders, and -.67 (*SEM* = .12) for sixth graders. Thus, fourth and sixth graders both used their monitoring to guide decisions about which texts to reread. The accurate monitoring for sixth graders would lead to effective regulation and theoretically better overall comprehension (as in Thiede et al., 2003). By contrast, even though fourth graders did not accurately monitor comprehension of texts after generating keywords, they used the results of the inaccurate monitoring to guide regulation of study. This would lead to ineffective regulation of study and likely worse overall comprehension than had they accurately monitored their comprehension.

Insert Figure 4 about here

General Discussion

In recent years, research has shown an increased interest in metacognitive monitoring and regulation of learning in children (Lockl et al., 2000; Lockl & Schneider, 2003; Vise et al., 2000). Typically, these studies show that, for simple learning material as word pairs, children are fairly well able to monitor their own learning as of age 8, and from the age of 10 onwards, they regulate their study behavior effectively. The question remains to what extent children are able to monitor and regulate learning of more complex, and educationally more relevant materials such as expository texts. This is the first study to examine the development of metacomprehension in elementary and middle school children. The present data suggest that as of grade 6, children are able to monitor comprehension of textual material when generating keywords prior to providing comprehension ratings. The poor metacomprehension accuracy in fourth graders suggests that the ability to judge comprehension of texts has not developed fully by age 10. It is important to notice, however, that we did not provide students with an extensive instruction on how to generate keywords and how to monitor comprehension. It is possible that even fourth graders would be able to benefit from generating keywords after more extended instruction and practice. The data on regulation of study reveal that, after generating keywords, even fourth graders were able to use their own comprehension ratings to decide which texts needed restudying. The gamma correlation of -.67 between comprehension ratings and text selections in sixth graders comes close to those previously found in adults (-.79, Thiede et al., 2003). Given the high gamma correlations between comprehension ratings and text selections, these data indicate that the regulation task was easier for students than the monitoring task. For fourth graders, this discrepancy between monitoring and regulation led students to select the wrong texts on a number of occasions: because students were unable to judge how well they had understood the texts, they chose the wrong texts for restudy, despite the high regulation gamma.

These data are surprising for a number of reasons. Previous studies have shown that the effective use of metacognitive skills during text reading does not develop until adolescence (Peverly, Brobst, & Morris, 2002), and that even college students have great difficulty self-regulating their learning when studying texts (Peverly, Brobst, Graham, & Shaw, 2003). In a study on fifth to twelfth graders and college students, Brown, Smiley, and Lawton (1978) found that only the college students adapted their learning of text material based on information from previous learning trials. Then again, college students working in the same paradigm as in the present study quite effectively regulated their study, when their monitoring accuracy was high (Thiede et al., 2003). Yet, the present data show that as of age 10, students can regulate study of texts based on self-generated comprehension ratings, and as of age 12, they are able to monitor their understanding of text fairly accurately. Why does generating keywords improve metacomprehension accuracy and regulation of study even in elementary school students? A number of factors seem to play a role.

First, previous research has shown that, in adults, summarizing text prior to rating comprehension stimulated students to base their ratings on gist information from the text (rather than detail information), which led to higher metacomprehension accuracy (Anderson & Thiede, 2008; Thiede et al., 2003). That is, summarizing a text at a delay after text study activates the situation model of the text, which is a good predictor of test performance (McNamara et al., 1996). The findings of studies on the effect of summarization mimic those of the keyword studies (Thiede et al., 2003, 2005), including the present ones. Generating keywords at a delay after text study may activate the situation model, which the student uses to assess his level of text comprehension. Note that we did not test their situation model directly and therefore can not draw any definite conclusions with regard to this explanation. Alternatively, rehearsal of the text information when generating keywords might have produced a memory effect, thereby increasing the chance of retrieval during judgments and improving the quality of the judgments. At this point, this explanation is less likely, as previous research has shown (Thiede et al., 2003, 2005) that the immediate keyword instruction did not improve metacomprehension accuracy. Moreover, in the present experiment keyword generation did not have an effect on test performance, which would have been expected in case of a memory effect.

The situation model explanation, however, does not answer the question why younger students fail to benefit from keyword generation. Research has shown that around fifth grade, students become aware of the structure of a text (Brown & Smiley, 1978) and learn to indicate important idea units in the text. The extent to which keywords activate the situation model depends on the quality of the generated keywords. As fourth grade students may be unable to assess the quality of their keywords, these keywords are an inaccurate basis for comprehension ratings. Thus, even though students at the age of 6 years old can use simple memory retrieval cues to monitor their learning, as the delayed-JOL effect in children indicates (Schneider et al., 2000), the ability to use more complex cues based on keyword generation does not develop until around the age of 11. This explanation is underlined by fourth graders effective regulation of study, which does not depend on keyword generation. An implication derived from this finding is that practicing generating keywords in class may not only help students to better identify the structure of a text, but may also improve their metacomprehension accuracy. Note that it is unlikely that text complexity caused lower metacomprehension accuracy in fourth than in sixth graders: the correlation between metacomprehension accuracy and test performance was –.04 (*p* = .59).

These findings shed new light on elementary school students' ability to monitor and regulate study of text materials. Whereas it was previously thought that metacognitive activities during text study are fairly inaccurate until well into adolescence, the present findings suggest otherwise. A relatively simple instruction as keyword generation induced accurate regulation of study in fourth to seventh graders, and improved metacomprehension accuracy in sixth and seventh graders. According to Brown's classification (1980) of metacomprehension, keyword generation improved metacomprehension and regulation on the first three of four levels: (1) knowing when you know and when you don't know, (2) knowing what it is you know, and (3) knowing what it is you need to know. Further research is needed to determine to what extent elementary and middle school children are able to monitor learning of texts at the fourth and most complex level of metacomprehension (knowing the usefulness of intervention strategies).

References

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

Ariel**,** R.**,** Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138,* 432-447.

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453-479). Hillsdale, NJ: Erlbaum.

Brown, A. L., & Smiley, S. S. (1978). The development of strategies for studying texts. *Child Development, 49,* 1076-1088.

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

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

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

Dunlosky, J., & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16,* 228-232.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition.* Beverly Hills, CA: Sage.

Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language, 36,* 34-49.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 702–718.

Goodman, L.A. & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association, 49,* 732-764.

Gravetter, F. J. & Wallnau, L. B. (1999). *Essentials of statistics for the behavioral sciences* (3rd Ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Hacker, D. J., Dunlosky, J., & Graesser, A.C. (Eds.) (2009). *Handbook of Metacognition and Self-Regulated Learning*. New York: Routledge.

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

Kobasigawa, A., & Metcalf-Haggart, A. (1993). Spontaneous allocation of study time in first- and third-grade children in a simple memory task. *Journal of Genetic Psychology, 154,* 223-235.

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

Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy: Insights from the processes underlying judgments of learning in children. In P. Chambres, M. Izaute, & P-J. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 1- 17). New York, NY: Kluwer.

Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology, 103,* 152-166.

Lockl, K., & Schneider, W. (2003). Metakognitive Überwachungs- und Kontrollprozesse bei der Lernzeiteinteilung von Kindern [Monitoring and control processes in children's allocation of study time]. *Zeitschrift für Pädagogische Psychologie, 17,* 173-183.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study*. Journal of Experimental Psychology: Learning, Memory, & Cognition, 10,* 756-766.

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

Masur, E. F., McIntyre, C. W., & Flavell, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrial free recall task. *Journal of Experimental Child Psychology, 15,* 237-246.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131,* 349-363.

McNamara, D. Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interaction of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14,* 1-42.

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

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

Nelson, T. O., Dunlosky, J., Graf, E. A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, *5*, 207-213.

Nelson, T.O., & Leonesio, R.J. (1988). Allocation of self-paced study time and the 'labor-in-vain effect'. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 476-486.

Nelson, T. O., & Narens, L. (1990). Metamemory: a theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-171). New York: Academic Press.

Peverly, S. T., Brobst, K. E., & Morris, K. S. (2002). The contribution of reading comprehension ability and meta-cognitive control to the development of studying in adolescence. *Journal of Research in Reading, 25,* 203-216.

Peverly, S. T., Brobst, K. E., Graham, M., & Shaw, R. (2003). College adults are not good at self-regulation: A study on the relationship of self-regulation, note taking, and test taking. *Journal of Educational Psychology, 95,* 335-346.

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

Roebers, C. M., von der Linden, N., & Howie, P. (2007). Favourable and unfavourable conditions for children's confidence judgments. *British Journal of Developmental Psychology, 25,* 109-134.

Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task. *Cognitive Development, 15*, 115–134.

Son. L. K. (2005). Metacognitive control: Children's short-term versus long-term study strategies. *The Journal of General Psychology, 132*, 347-363.

Son, L. K. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 255-262.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science, 3,* 315-316.

Thiede, K. W. (1999). The importance of accurate monitoring and effective self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6,* 662-667.

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28,* 129-160.

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

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

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

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D.J. Hacker, J. Dunlosky, & A.C. Graesser, (Eds.) *Handbook of metacognition and self-regulated learning* (pp. 85-106). Mahwah, NJ: Erlbaum.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology, 132,* 408-428.

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

Table 1.

Test Performance and Comprehension Ratings

Group	Test Performance	Comprehension Rating
	Experiment 1	
Seventh Grade		
No-Keyword	2.58 (.15)	2.90 (.15)
Delayed-Keyword	2.54 (.15)	3.02 (.13)
	Experiment 2	
Fourth Grade		
No-Keyword	1.92 (.16)	3.47 (.13)
Delayed-Keyword	2.09 (.15)	2.98 (.12)
Sixth Grade		
No-Keyword	2.36 (.17)	3.17 (.14)
Delayed-Keyword	2.65 (.17)	3.21 (.14)

Note. Entries are mean across individual's median test performance and median comprehension ratings. Values in parentheses are standard errors of the means. Both test performance and ratings are on a 5-point scale.

Appendix A

Cellular Respiration

You've been hiking all morning, and you are hungry. You get out a sandwich you packed and begin munching. Food supplies your body with glucose, an energy-rich sugar. Respiration is the process by which cells obtain energy from glucose. During cellular respiration, cells break down simple food molecules such as glucose and release the energy they contain. Most of the energy used by the cells in your body is provided by cellular respiration.

Energy stored in cells is something like money in a savings account. During photosynthesis, plants capture energy from sunlight and "save" it in the form of carbohydrates, including sugars and starches. When you eat, you add to your body's energy savings account. When cells need energy, they "withdraw" it by breaking down the carbohydrates in the process of respiration.

Cellular respiration is a two-stage process. The first stage takes place in the cytoplasm of the organism's cells. There, molecules of glucose are broken down into smaller molecules. Oxygen is not involved, and only a small amount of energy is released.

The second stage of cellular respiration takes place in the mitochondria. There, the small molecules are broken down into even smaller molecules. These chemical reactions require oxygen, and they release a great deal of energy. This is why the mitochondria are sometimes called the "powerhouses" of the cell.

Energy is released as a product in both stages of respiration. This is transferred to other molecules, which then carry the energy where it is needed for the activities of the cell. The rest of the energy is released as heat. Two other products of cellular respiration are carbon dioxide and water. These products diffuse out of the cell. In most animals, the carbon dioxide and some water leave the body during exhalation, or breathing out. When you breathe in, you take in oxygen – a raw material for respiration. When you breathe out, you release carbon dioxide and water.

Questions

Considering the comparison of respiration to a savings account, a fatigued muscle would be analogous to
- a. money
- b. a bank director
- c. a person living in poverty
- d. an armored van for transporting money

When you breathe into your hand and your palm becomes moist, this is caused by:
- a. The release of oxygen through your skin
- b. Your breath being warmer than the air around it
- c. Water from your cells leaving your body in your breath
- d. A normal breakdown of the cells' cytoplasm

If glucose does not break down in your cells, you will feel,
- a. Tired
- b. Out of breath
- c. Hyperactive
- d. Overheated

Why would a runner take off her jacket on a cold day?

a. She ate too much glucose before the run
b. Carbon dioxide combines with water to produce heat as she runs
c. As glucose breaks down, the runner's cells release both energy and heat
d. Starches combine with sugars to make her sweat

Which of the following would improve an athlete's performance in a race?

a. Low carbohydrate snack, low oxygen environment
b. Low carbohydrate snack, high oxygen environment
c. High carbohydrate snack, high oxygen environment
d. High carbohydrate snack, low oxygen environment

Appendix B

Elephants

Elephants are the largest and heaviest terrestrial animals. An African male elephant (bull) weighs about 80 people, 6 cars, 12 large horses or 1200 cats! Elephants are exceptionally strong and can lift entire trees with their trunk. They are also very intelligent and friendly animals. The females live together in family groups and take care of each other. Just as humans, elephants are mammals. Like all mammals elephants regulate their own body temperature. Elephants are, after humans, the longest living mammals: they can live to up to 70 years. There are two sorts of elephants: The African and the Asian elephant. Both sorts have a long trunk, big ears, and a thick, grey skin.

How can you tell and African and an Asian elephant apart? They look like each other, but there are differences. The clearest difference is found in the size of their ears: those of the African elephant are larger. African elephants have longer legs and a thinner body as their Asian counterparts. The back of the Asian elephant is curved up, the back of the African elephant is curved down. Another difference concerns their tusks: In Asia only the males usually have tusks, in Africa both the males and females usually have tusks.

The two kinds of elephants are subdivided in smaller groups we call subspecies. These subspecies all have a somewhat different appearance and are named after the territory they live in. Three subspecies live in Africa: the savanna elephant living on open grass, the bush elephant of Western and Central Africa and the desert elephant in Namibia. The most important subspecies in Asia are the Indian elephant and the Sumatran elephant that lives on the Indonesian islands Sumatra and Borneo. T he Sumatran elephant is the smallest of the three subspecies, it is also the lightest in color and has less pink spots than the other Asian subspecies.

Questions

What is the same in the Asian and African elephant?

a. Back

b. Tusks

c. Trunk

d. Ears

What is said in the text about elephants?

a. Elephants are the most friendly terrestrial animals

b. Elephants are the strongest terrestrial animals

c. Elephants are the heaviest terrestrial animals

d. Elephants are the tallest terrestrial animals

What is said in the text about tusks?

a. Only Asian elephants have tusks

b. Only African elephants have tusks

c. Only Asian females and African elephants have tusks

d. Only Asian males and African elephants have tusks

What is the difference between Asian and African elephants?

a. Asian elephants have shorter legs

b. African elephants are heavier

c. African elephants have smaller ears than Asian elephants

d. African elephants are smarter

Author Note

We would like to thank Eleni Pons, Katinka Beker, and Sophie Kapiteijn for their help in conducting the experiments and analyzing the data. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B07460 to Keith W. Thiede, Thomas D. Griffin, and Jennifer Wiley. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Footnotes

1. Nelson (1984) recommended using a Goodman-Kruskal gamma correlation (Goodman & Kruskal, 1954) for these kinds of data. Gamma is computed by examining the direction of one variable relative to another. If one variable (e.g., metacomprehension judgment) is increasing from one text to another and the other variable (e.g., test performance) is also increasing across this same pair of texts, this is considered a concordance (C). By contrast, if one variable is increasing from one text to another and the other variable is decreasing across this same pair of texts, this is considered a discordance (D). Concordance and discordance is computed across all pairs of items. The total number of each is used to compute the correlation coefficient, Gamma = (C – D)/(C + D).

Figure Captions

Figure 1. Mean metacomprehension accuracy for seventh grade students (Experiment 1).
Figure 2. Mean regulation of study for seventh grade students (Experiment 1).
Figure 3. Mean metacomprehension accuracy for fourth and sixth grade students (Experiment 2).
Figure 4. Mean regulation of study for fourth and sixth grade students (Experiment 2).