READING ABILITY AND SUMMARIZATION EFFECTS ON METACOMPREHENSION ACCURACY

A Thesis

Presented in Partial Fulfillment of the Requirements for

the Degree Master of Arts in the

Graduate School of The Ohio State University

By

Jacqueline Janice von Spiegel, B.A.

*****

The Ohio State University
2006

Master's Examination Committee:

Dr. Dorothy Jackson, Adviser

Dr. Sarah T. Boysen

Dr. Gerald Winer

Approved by

[Signature]

Adviser

Graduate Program in Psychology
ABSTRACT

In the present study, the factors influencing accuracy of comprehension monitoring (metacomprehension accuracy) were investigated. College students first completed a standardized test of reading ability. Next, they read three texts; then some wrote a summary of each text (either immediately after reading or after a delay), whereas others did not write summaries. All the participants then rated their comprehension of each text. Finally, they completed a comprehension test over the material covered in each text. Metacomprehension accuracy, the correlation between ratings of comprehension and later test performance, was affected by both summarization and reading ability. This study also examined the mechanisms behind the summarization and reading ability effects by investigating the features of the summaries themselves.
Dedicated to

Brandon,

For keeping me sane when I needed to be, and bringing blissful insanity to my life

The rest of the time.

And to

Aidan Robinson,

My nephew and godson, who was born on the same day as this paper was

completed.
ACKNOWLEDGMENTS

First, I would like to express great thanks to my adviser, Dr. Dorothy Jackson, for her guidance, support, and enthusiasm throughout the entire thesis process.

Also, I wish to express many thanks to the undergraduates who helped me with various aspects of my research during the past few quarters. Many thanks to Ashley Jones for the meticulous data collection and to Jaclyn Marnell, Scott Sutton, Jeneen Thrash, and Christina Ellinger for their very thorough data analysis.

Many thanks as well to Dr. Laura Wagner and Ellen Furlong for their time and invaluable advice.

Finally, I would like to thank my family and friends, for believing in me and for patiently waiting while I took time away from them to work on this project. I would especially like to thank Charley and Jan von Spiegel, for their endless support and love.
VITA

August, 2001 .................................................. B.A. Psychology, Cum Laude
University of Akron

2004 ........................................................... "Automatic processing of
elements interferes with
processing of relations"
Paper Presentation (with Dr.
Vladimir Sloutsky),
Cognitive Science Society
Conference

2002 – Present .............................................. Departmental Fellowship,
Psychology Department,
The Ohio State University

FIELDS OF STUDY

Major Field: Psychology
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Literature Review</td>
<td>2</td>
</tr>
<tr>
<td>1.1.1 Metacognition</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2 Components of Metacognition</td>
<td>2</td>
</tr>
<tr>
<td>1.1.3 Metacognitive Strategy Development</td>
<td>4</td>
</tr>
<tr>
<td>1.1.4 Monitoring Accuracy</td>
<td>5</td>
</tr>
<tr>
<td>1.1.5 Metacomprehension Accuracy</td>
<td>6</td>
</tr>
<tr>
<td>1.1.6 Discrepancy-Reduction Models</td>
<td>7</td>
</tr>
<tr>
<td>1.1.7 Reading Ability and Metacomprehension Accuracy</td>
<td>8</td>
</tr>
<tr>
<td>1.1.8 Factors that can Increase Metacomprehension Accuracy</td>
<td>11</td>
</tr>
<tr>
<td>1.2 Summary and Hypotheses</td>
<td>13</td>
</tr>
<tr>
<td>2. Methods</td>
<td>15</td>
</tr>
<tr>
<td>2.1 Participants</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Materials</td>
<td>16</td>
</tr>
<tr>
<td>2.3 Procedure</td>
<td>17</td>
</tr>
<tr>
<td>3. Results</td>
<td>21</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Grouping of participants by condition and reading ability</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>Detailed description of procedure for each condition</td>
<td>20</td>
</tr>
<tr>
<td>3.1</td>
<td>Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each reading group level</td>
<td>23</td>
</tr>
<tr>
<td>3.2</td>
<td>Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each condition</td>
<td>24</td>
</tr>
<tr>
<td>3.3</td>
<td>Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each reading group level and condition</td>
<td>25</td>
</tr>
<tr>
<td>3.4</td>
<td>Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each text</td>
<td>26</td>
</tr>
<tr>
<td>3.5</td>
<td>Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each reading group level (without <em>Colosseum</em>)</td>
<td>26</td>
</tr>
<tr>
<td>3.6</td>
<td>Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each condition (without <em>Colosseum</em>)</td>
<td>27</td>
</tr>
<tr>
<td>3.7</td>
<td>Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each reading group level and condition (without <em>Colosseum</em>)</td>
<td>27</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.1</td>
<td>Comparison of the means (and standard errors) of the number of details for better and poor readers in the Delayed and Immediate Summary conditions</td>
<td>30</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

One aspect of cognition that is becoming more central in the study of cognitive development is metacognition. Metacognition is a construct that involves an awareness and management of one's own cognitive processes, or "thinking about thinking."

Although metacognitive abilities are evidenced early in life, when children become aware of their own and other's thoughts, these skills continue to develop through adulthood. In addition, the management of one's own cognition is a crucial component in succeeding in many academic and social situations. For example, in debates with someone holding an opposing viewpoint, young adolescents are readily able to provide support for an argument, but have difficulty in attending to their opponent's viewpoint and addressing their opponent's supporting argument through counterargument (Felton & Kuhn, 2001).

Since these metacognitive skills are unlikely to develop in the absence of suitable educational environments, psychological researchers need to examine the mechanisms behind the development of these skills in order to gain a fuller understanding of metacognition. The present study seeks to further this line of research by assessing the metacognitive abilities of college students and subsequently examining the factors that contribute to the strength of those abilities.
1.1: LITERATURE REVIEW

1.1.1: METACOGNITION

Metacognition, as defined by Flavell (Flavell, Miller, & Miller, 2002), is "any knowledge or cognitive activity that takes as its object, or regulates, any aspect of any cognitive enterprise" (p. 164). In other words, metacognition can be described as "thinking about thinking." Metacognition, sometimes referred to as "executive control," is crucial in cognitive development (Kuhn & Dean, 2004). Children not only think about the task at hand when they are solving a problem, but they also learn to think about thinking and about strategies and the process of solving a problem. Metacognitive skills play an important role in many types of cognitive activities that are related to problem solving, including reading comprehension, writing, self-instruction, and logical reasoning, to name just a few. Metacognition can be related to Piagetian formal operational thinking because formal operational thinking involves thinking about propositions, hypotheses, and imagined possibilities, which are all forms of cognition (Flavell, Miller, & Miller, 2002). In sum, becoming aware of and reflective about one’s own thinking is a vital part of successful problem solving and learning. Hence, the study of the development of metacognition is essential to the further understanding of children’s thinking.

However, before we delve into the development of metacognitive skills, it is important to discuss the distinctions made among the various aspects of metacognition.

1.1.2: COMPONENTS OF METACOGNITION

Metacognition is thought to include two components: knowledge about cognition and regulation of cognition (Baker & Brown, 1984). A person’s knowledge about
cognition refers to the general information and stable knowledge that he or she possesses concerning cognitive processes and his or her awareness about strategy use when engaged in such processes (e.g., activities such as reading). For example, Jackson, Smith, and Paul (1997) found that probing for prior knowledge on a topic, and thereby activating metacognitive awareness of strategies, increased the subjects’ reading comprehension. This finding shows the importance of knowing the availability of strategies and the effective usage of strategies (i.e., knowledge about cognition) in becoming a better learner.

Regulation of cognition, on the other hand, refers to the effectiveness with which learners keep track of ongoing cognitive processes and their employment of regulatory strategies in order to solve problems. In other words, regulation of cognition is how learners control their learning.

As noted by Baker and Brown (1984), the two metacognitive components (knowledge about cognition and regulation of cognition) are “somewhat separate phenomena” (p. 353). Knowledge about cognition is a relatively stable component. An individual who knows general facts regarding text processing and comprehension strategies is likely to continue to know these facts when asked about them. However, activities related to regulation of cognition, such as checking comprehension and implementing useful strategies, are not necessarily stable skills. For example, a reader who is aware of the effectiveness of certain strategies may not actually apply them in certain situations due to a heavy cognitive load or a lack of interest or motivation. Thus, although one’s knowledge of cognition may frequently enhance one’s regulation of cognition, the former does not guarantee the latter. Therefore, study of the skills
involved in the regulation of cognition is essential in understanding the development of metacognition in learners.

1.1.3: METACOGNITIVE STRATEGY DEVELOPMENT

Extant research on metacognition has shown that regulatory abilities develop throughout childhood and adolescence, and sometimes are not even achieved until adulthood. Research on the development of regulation of cognition has focused on the use of strategies, which are defined as cognitive or behavioral activities that a person may voluntarily carry out in order to enhance performance (Flavell, Miller, & Miller, 2002; Siegler, 1998). Strategy use is central to the study of the development of regulation of cognition because regulating involves selecting appropriate strategies for use, deciding on when to change strategies, and monitoring strategy success (Schraw, 1998).

Consequently, the current developmental model of regulation of cognition concentrates on the production and use of strategies (Flavell, Miller, & Miller, 2002; Schneider & Sodian, 1997). This model focuses on learners' abilities and deficiencies in metacognitive strategy use. Metacognition and strategy use appear to form a feedback loop in which feedback about the success or failure of each strategy increases metacognitive awareness of the goal and the value of each strategy in meeting that goal, and learners use this feedback to progress through the following stages (Miller, 2002).

Initially, children do not possess any of the skills necessary to carry out any strategies. Since the child at this stage does not have the ability to execute the strategies, any attempts to elicit the use of a strategy will be ineffective. As children develop the cognitive skills necessary to use the strategies, they will transition into the mediational deficiency stage (Schneider & Sodian, 1997). In this stage, children are capable of using
strategies, but any attempts to execute the strategy (either spontaneously or elicited) are not fruitful. As children become more familiar with these strategies, they progress into the production deficiency stage. Here, children do not produce strategies spontaneously, but elicited strategies are beneficial to performance. Next, children attempt to spontaneously produce strategies. In this stage, called utilization deficiency, the strategies that children spontaneously produce provide no benefit, or less benefit than for older children. The last period is characterized by reasonably mature and sophisticated use of appropriate strategies, becoming even more fine-tuned during adolescence and early adulthood.

This model describes the developmental progression of metacognitive strategy use, but it noticeably does not indicate specific age ranges at which the transitions take place. Children, adolescents, and even adults show these deficiencies on a wide range of strategies and tasks, and over a wide age range (Bjorklund, Miller, Coyle, & Slawinski, 1997). Due to the domain-specific nature of the development, the ages of transition are not as significant as the path the development takes.

1.1.4: MONITORING ACCURACY

Another one of the skills involved in regulation of cognition is monitoring (Schraw, 1998). Monitoring refers to a learner's awareness of comprehension, task performance, and strategy use. A common way of testing monitoring ability is by assessing monitoring accuracy. In monitoring accuracy, a person studies an item, then judges how well he or she learned the item, and finally is tested on the item. Monitoring accuracy refers to the correlation between the judgment of learning and performance on
the test. In other words, participants demonstrate accurate monitoring if they "know what they know and know what they do not know."

1.1.5: METACOMPREHENSION ACCURACY

The majority of studies investigating monitoring accuracy use associated learning tasks (i.e., learning word pairs). However, a more applied approach is to use texts as study items instead of pairs of words. Using texts increases the ecological validity of the experiment by using study items that could actually be used in classrooms. When using texts to examine monitoring accuracy, the paradigm is called metacomprehension accuracy. This paradigm, analogous to monitoring accuracy paradigms, considers the accuracy with which readers evaluate the end state of their comprehension (Glenberg & Epstein, 1985, 1987; Maki & Berry, 1984; Weaver, 1990). Thus, metacomprehension accuracy not only involves reading for comprehension (a cognitive activity) but also the processes of knowing when comprehension has been attained (a metacognitive activity). In this paradigm, participants read a passage and then are asked to make predictions about their future performance on a comprehension test. For example, readers may be asked to rate their confidence in their ability to answer questions about the passage just read, to predict how well they expect to do on a comprehension test, or to assess how well they think they understand the passage on a Likert-type scale (e.g., Glenberg & Epstein, 1987; Maki, Foley, Kajer, Thompson, & et al., 1990; Thiede & Anderson, 2003; Weaver & Bryant, 1995). Metacomprehension accuracy is operationalized as the relation between confidence and performance or, alternatively, between predicted and actual performance.
1.1.6: DISCREPANCY-REDUCTION MODELS

Metacomprehension accuracy is essential to optimal learning because it allows
the reader to be able to distinguish between well-learned material and less-learned
material. This distinction would further allow the reader to more effectively choose
material for study, allocating more time for less-learned material so as to optimize
learning for all material. Discrepancy-reduction models of self-regulated study (Butler &
Winne, 1995; Dunlosky & Hertzog, 1997; Dunlosky & Thiede, 1998; Koriat &
Goldsmith, 1996; Nelson & Narens, 1994; Thiede & Dunlosky, 1999) posit that learners
monitor their understanding of the material while studying, and the results of their
monitoring determine whether further study is required. That is, if the current state of
learning is less than the desired state of learning, the learner will continue to study. This
model suggests that accurate monitoring of comprehension is necessary for effective
regulation of study, which in turn, will optimize overall learning (Son & Metcalfe, 2000).

Thiede (1999) found support for these claims by analyzing which items participants
selected for restudy (see also de Bruin, Rikers, & Schmidt, 2005; Sperling, Howard,
Staley, & DuBois, 2004; Thiede, Anderson, & Therriault, 2003; Thiede & Dunlosky,
1999). After judging their learning on a paired associate task (i.e., learning word pairs),
participants were asked to select certain items for restudy. It was found that participants
who were more accurate at monitoring their learning selected for restudy the items they
considered most difficult to learn. This demonstrates effective self-regulation of study
time. In addition, the participants who were more effective in their choice of items for
restudy also obtained higher scores on recall tests. These findings provide support for the
discrepancy-reduction model by indicating that the more accurately a person monitors
comprehension, the more effective he or she is at regulating study time, and consequently, the better he or she performs on comprehension tests of the material.

1.1.7: READING ABILITY AND METACOMPREHENSION ACCURACY

Studies of metacomprehension accuracy have found a positive correlation between reading ability and monitoring accuracy (Castenell, 1983; Glover, 1989; Maki & Berry, 1984; Shaughnessy, 1979). As noted by Pressley (1995), skilled readers tend to engage in many conscious cognitive processes to extract meaning from texts. They are more likely to use different strategies (e.g., focusing on relevant information and relating important ideas to one another) and actively evaluate and monitor reading comprehension. Poor readers tend to display production or utilization deficiencies in metacognitive strategy use.

Previous research provides evidence that there is a relation between reading ability and post-test estimations of comprehension (Castenell, 1983; Shaughnessy, 1979). That is, better readers were more accurate at predicting their performance on a comprehension test after taking it than less proficient readers were. Researchers have also examined reading ability as a factor influencing an individual’s ability to predict comprehension performance before taking the comprehension test (i.e., metacomprehension accuracy). Maki and Berry (1984), in the first experiment described in the paper, divided their participants into those scoring above and below the median test performance in order to investigate the metacomprehension accuracy of better versus poorer students. Undergraduate students read chapters from introductory psychology texts dealing with unfamiliar topics and rated their confidence in their ability to answer multiple-choice questions based on each section of the chapter after a 1-day delay. The
participants had poor metacomprehension accuracy and were generally overconfident in their ability to answer the comprehension questions. However, when the sample was divided by performance on the comprehension tests, there was a difference in metacomprehension accuracy (measured by Pearson’s $r$ correlation coefficient). Those who scored above the median on the tests showed a small amount of monitoring accuracy ($r = .15$), but those below the median did not ($r = -.03$). This finding, that those who performed better on the comprehension test (i.e., better readers) were also more accurate at monitoring their comprehension, demonstrates the positive relation between reading ability and metacomprehension accuracy.

Glover (1989) also compared good and poor readers’ metacomprehension accuracy, but used a more standard measure of reading ability in college students instead of using the scores on the comprehension tests as in the previous research. In the study, over two hundred undergraduate students were given the Nelson–Denny Reading Test (Brown, Bennett, & Hanna, 1981), and the 15 students with the highest scores and the 15 with the lowest scores were identified. The results indicated that overall correlations between students’ estimates of comprehension and their actual test performance were close to zero ($r = .04$). Instead, the difference between each student’s estimated and actual test performance was used as a measure of monitoring accuracy. The lower the difference score, the more accurate the performance estimations, and therefore the more accurate the metacomprehension accuracy. When good and poor readers’ difference scores were compared, the difference scores of good readers were significantly lower ($M = 1.21$) than those of poor readers ($M = 7.43$), indicating that good readers exhibited a much higher accuracy of comprehension monitoring than poor readers. Poor readers
were not only inaccurate in estimating their future test performance but also often wildly overconfident in their future performance.

In order to test these findings in a more ecologically valid setting, this model was taken out of the laboratory and into the classroom (Hacker, Bol, Horgan, & Rakow, 2000). Hacker and colleagues investigated the relation of higher and lower performing students with their metacomprehension accuracy on exams in an existent college course. Students were given practice exams and corresponding answer keys one week before their actual exams, and they were encouraged to use these practice exams to determine their strengths and weaknesses in their understanding of the material. Immediately before the actual exam one week later, the students were asked to give a prediction of their performance on the exam and their estimated study times. Students were divided into a high performance group and a low performance group based on their exams scores being above or below the median. Low performance students demonstrated poor metacomprehension accuracy on the first exam and showed no improvement by the end of the course. However, although the high performance students also demonstrated poor metacomprehension accuracy on the first exam, they showed increases in metacomprehension accuracy by the end of the course. Low performance students showed strong overconfidence in their predictions of their exam scores, becoming more exaggerated as the scores decreased. This research portrays a bleak future for the lowest performing students. Not only do they fail to comprehend the course content, but they demonstrate a lack of awareness of their own deficits and show no improvements in this awareness, even with repeated feedback. Also, prior performance on early exams did not have an impact on students' subsequent study times. This could be due to the students'
failure to use their self-monitoring to adjust their behavior accordingly. These results provide evidence for the aforementioned disconnection of the knowledge about one's cognition and the regulation of one's cognition. It also demonstrates how, on certain tasks, even adolescents and adults can demonstrate production and utilization deficiencies in metacognitive strategies.

Although research on the effect of reading ability on metacomprension accuracy is sparse, evidence suggests that the effect is positive. More skilled readers, in most cases, are more accurate in monitoring their comprehension than less skilled readers.

1.1.8: FACTORS THAT CAN INCREASE METACOMPREHENSION ACCURACY

Researchers studying readers' ability to monitor comprehension have not only verified the correlates of metacomprension accuracy, but they have also begun to investigate factors that may increase the accuracy of monitoring. Early investigations of metacomprension accuracy conducted by Glenberg, Epstein, and their colleagues (Glenberg & Epstein, 1985, 1987; Glenberg, Wilkinson, & Epstein, 1982) suggested that undergraduate participants were very poor at accurately monitoring text comprehension, often overestimating and sometimes underestimating how much had been comprehended. Confidence ratings, in fact, were often just barely correlated with performance scores (i.e., correlations were never higher than 0.20).

Although early studies indicated that readers were not able to accurately monitor comprehension, more recent studies have shown reliable (although still low) correlations between test predictions and test performance. In previous research, investigators often
used a single test item per text to measure comprehension (Glenberg & Epstein, 1985, 1987). Although the reading materials employed were short passages, with each being one-paragraph in length, using a single item per text as a measure of comprehension seriously constrained the reliability and validity of comprehension measurement. Clearly, comprehension is a continuous variable and should be measured by multiple questions. Several studies (Maki et al., 1990; Maki & Serra, 1992a, 1992b; Weaver, 1990) tried to modify the methodology used to assess metacomprehension accuracy. By asking undergraduate readers multiple questions about each passage, as opposed to just one, these researchers have demonstrated that previous research using a single inference per text as a measure of comprehension often failed to reveal the readers’ actual level of metacomprehension accuracy because comprehension ability is rarely an “either-or” phenomenon. These studies have consistently detected effects showing that undergraduate participants could at least accurately monitor their comprehension at a moderate level (i.e., mean correlations of about 0.35).

Along with varying the comprehension tests and passages for reading, researchers also vary the timing of the judgment of comprehension. Nelson and Dunlosky (1991) found a *delayed JOL* (judgment of learning) *effect* when studying the accuracy of participants’ predictions of test performance. Their method was to delay the prompt for assessing metacomprehension for a short while (during which other items are studied). This delay is crucial in enabling participants to clear their working memory (WM) and access only items stored in long-term memory (LTM) when making a judgment of how well they know the material. This delay enables participants to clear their WMs of relevent information and, thereby, more clearly evaluate the information stored in LTM.
The results of this simple manipulation in previous research has been more accurate metacomprehension ratings, which in turn increases the efficacy of self-regulated studying and results in higher test performance (Nelson & Dunlosky, 1991; Thiede & Anderson, 2003; Thiede, Anderson, & Therriault, 2003; Thiede & Dunlosky, 1994; Thiede, Dunlosky, Griffin, & Wiley, 2005).

In addition to modifying the methods of assessment, researchers have found interventions that can improve readers’ metacomprehension accuracy. One such intervention is to instruct the readers to write summaries of the texts before judging their comprehension. In a study by Thiede and Anderson (2003), participants read expository texts; some participants summarized these texts (either immediately after reading each text or after a delay, which was filled by reading the remaining texts) and others did not summarize (control group). After summarizing, all participants rated their comprehension of each text and answered comprehension questions over the content of the texts. Metacomprehension accuracy was found to be greater when summaries were written after a delay than when summaries were not written or were written immediately after reading. This suggests that having readers summarize texts can improve their metacomprehension accuracy, but only if it is after a delay.

1.2: SUMMARY AND HYPOTHESES

Identifying the factors that influence metacomprehension accuracy provides a means of investigating the cognitive and metacognitive processes involved in comprehension monitoring. In previous literature, we find summarization and introduction of a delay between study and judgments of learning. However, the summarization effect may not be true for all readers. Since previous research has found
that more capable readers already have better metacomprehension accuracy and are more capable of using metacognitive strategies to their benefit, the interventions of summarization may differentially improve readers' monitoring based on their current reading ability. This could be due to differences in prior strategy use, including the presence of production or utilization deficiencies in poorer readers. In the present study, the effects of summarization and delay of judgment are examined within the context of the learner's general reading ability. Specifically, three hypotheses are put forth. First, it is expected that better readers will be more accurate at monitoring their comprehension than poorer readers. Second, based on the findings of previous research, (Thiede & Anderson, 2003) it is predicted that summarizing the texts will improve metacomprehension accuracy, but only if there is a delay between study and judgment of learning. Finally, and most importantly, it is expected that the summarization intervention will differentially improve metacomprehension accuracy in better and poorer readers. That is, summarization after a delay will improve metacomprehension accuracy for both groups of readers, but there will be a greater significant difference in the size of the improvement for the poor readers, as opposed to the better readers.
CHAPTER 2

METHODS

In this quasi-experimental design, there were three independent between-subjects variables (i.e., reading ability, summarization, and delay of summarization). The effects of these variables were measured on one dependent variable (i.e., metacomprension accuracy). As shown in Table 2.1, participants were divided into six different groups. It is quasi-experimental because although the participants were randomly assigned into one of three conditions, they were not randomly assigned into reading ability groups. Participants were placed into reading ability groups based on their performance on a test of reading ability.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Reading ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>No summarization</td>
<td>Better</td>
</tr>
<tr>
<td></td>
<td>Group 1</td>
</tr>
<tr>
<td></td>
<td>Poorer</td>
</tr>
<tr>
<td></td>
<td>Group 2</td>
</tr>
<tr>
<td>Summarization without delay</td>
<td>Group 3</td>
</tr>
<tr>
<td></td>
<td>Group 4</td>
</tr>
<tr>
<td>Summarization with delay</td>
<td>Group 5</td>
</tr>
<tr>
<td></td>
<td>Group 6</td>
</tr>
</tbody>
</table>

Table 2.1: Grouping of participants by condition and reading ability.
2.1: PARTICIPANTS

The participants were introductory psychology undergraduate students at a large Midwestern university. The sample consisted of 101 participants (40 males, 61 females) mean age of 18.3 years. The ethnic makeup of the sample consisted of 87 Caucasians, 5 Asian-Americans, 1 Hispanic, 6 African-Americans, and 2 participants identified as “Other.” The participants were all freshman in their first quarter at the university. The participants received course credit for their participation.

2.2: MATERIALS

All materials used were the following:

1. Demographic Information. A demographic questionnaire was designed to collect relevant information (i.e., age, gender, and race). This information was not used in the analysis; it was only used for reporting the characteristics of the sample.

2. Nelson-Denny Reading Test. The Nelson-Denny Reading Test (Brown, Fishco, & Hanna, 1993) was used to determine reading skill level by testing reading comprehension. The Nelson-Denny Reading Test (NDRT) is a standardized and well-known measure of reading ability used for high school and college students. The NDRT is composed of two subtests: Vocabulary and Comprehension. The Vocabulary subtest consists of 80 items, each with five answer choices, and has a total time limit of 15 minutes. The Comprehension subtest consists of seven passages of text and a total of 38 questions, each with five answer choices. The total time limit is 20 minutes for this subtest. The participants’ scores are determined by summing the number of items answered correctly in each subtest.
3. **Texts.** Three passages that were used as texts were also used by Jelinek-Lewis and Jackson (2001) in a study on comprehension of television captioning. The passages are excerpts from the transcripts of narrated documentaries shown on television. These passages were used in the present study in order to replicate Thiede and Anderson’s (2003) results using a different type of text than is typically used in studies of metacomprehension accuracy. The texts are approximately 1100 words long and cover a variety of topics (i.e., Stonehenge, The Incas, and the Colosseum), and have Flesch–Kincaid readability scores of 8.1, 8.9, and 10.0, respectively.¹ A sample text is provided in Appendix A.

4. **Comprehension Tests.** Comprehension tests over the passages were also developed previously, including explicit and implicit types of questions (Jelinek-Lewis & Jackson, 2001). The comprehension tests were 18-question multiple-choice (with four answer choices). Jackson, Paul, & Smith (1997) found that these comprehension tests predict SAT scores very well ($p < .0001$). Sample questions are provided in Appendix A.

2.3: **PROCEDURE**

The procedure used in this study is modeled after the procedure used in Thiede and Anderson (2003) in order to replicate their findings and for the present results to be comparable to previous metacomprehension accuracy research. The participants were

---

¹ Readability scores may not be a true indicator of text difficulty. They take into consideration only the length of sentences and the number of syllables. In other words, they take into account only the actual text, not the content or the reader. Other important factors that determine difficulty are the topic of the text and the reader’s prior knowledge and/or interest in the topic (Schraw, 1998). Additionally, some texts may meet the requirement of “simpler” texts (they have shorter sentences with fewer multi-syllabic words in them), and yet are more difficult to understand because they have important elements for comprehension deleted in order to produce shorter sentences (Thiede & Anderson, 2003). Furthermore, many multi-syllabic words are familiar to most readers, yet because they are included in a text, they increase the readability score of the text.
assigned to one of three conditions based on order of appearance. The first condition (No Summary Control Condition) followed the standard metacomprehension accuracy procedure similar to that developed by Glenberg and Epstein (1985). Participants in this condition completed the NDRT and then read the three texts. After reading all the texts, they rated their comprehension of each text. The comprehension rating was measured by a Likert-type question format (i.e., “How well do you think you understood this passage?”) followed by the title of the passage and a scale of 1 to 7 (1 = very poorly). The passages were rated in the order that they were read. Then these participants completed the comprehension tests over the three texts (in the order that they were read). This procedure has been widely used in the metacomprehension accuracy literature (see Glenberg & Epstein, 1987; Maki & Serra, 1992a; Weaver, 1990 for a review). Hence, this control group was included to provide a link to the literature and to verify that a level of accuracy similar to that in previous research is demonstrated.

Participants in the second condition (Delayed Summary Condition) first completed the NDRT and then read the three texts. After reading all the texts, they were given a paper with the title of one text on it and asked to write a summary of that text. Summaries of all three texts were written in the order that they were read. After writing all three summaries, they rated their comprehension of each text, and then completed the comprehension tests.

Participants in the third condition (Immediate Summary Condition) first completed the NDRT. They then read the first text and immediately wrote a summary of it (without reference to the text). They read and immediately summarized each of the three texts. These participants then rated their comprehension of each text and completed
the comprehension tests. Based on previous findings regarding the timing of summarization, it is predicted that metacomprehension accuracy will be greater for the Delayed Summary Condition than for the Immediate Summary Condition or the No Summary Control Condition.

All participants were instructed that they would take a reading comprehension test, read texts, rate their comprehension of the texts, and then answer test questions about the texts. Participants were also instructed that they might also be asked to write summaries of the texts. All participants, including those in the No Summary Control Condition, were given paper and a pen. The order of presentation of the texts (and subsequent summaries, comprehension ratings, and comprehension tests) was randomized. An overview of the procedure is presented in Table 2.2.
Table 2.2: Detailed description of procedure for each condition.
CHAPTER 3

RESULTS

In previous studies (e.g., Glenberg & Epstein, 1987; Maki & Berry, 1984; Weaver, 1990), monitoring accuracy was operationalized as a Goodman–Kruskal gamma correlation between a participant’s comprehension rating and test performance across texts. However, because in the present study each participant read only three texts, the gamma correlation across texts for each individual was not an accurate representation of their metacomprehension accuracy. Therefore, the median comprehension rating and the median test performance for each individual was computed. Consequently, metacomprehension accuracy was operationalized as the correlation between these comprehension ratings and test performances across groups. The median was used 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 (Pagano, 1998). Throughout this investigation, Pearson correlation coefficients were computed along with gamma correlations as a measure of association. These measures of association led to convergent conclusions throughout this investigation. And because more operations can be performed on the Pearson correlation coefficient (e.g., hypothesis testing), only this statistic is reported.
For each of the following correlations, t tests were conducted to evaluate whether
the Pearson correlation for each group was significantly different than zero. The t test of
the null hypothesis for Pearson’s r, H₀: ρ = 0, is:

\[ T = r\sqrt{n - 2}/ \sqrt{1 - r^2}, \text{ with } df = n - 2. \]

Significance in these t tests indicates that the comprehension ratings were predictive of
subsequent test performance for participants in the group being tested. In the following
analysis, most of these correlations were significant, so for the sake of brevity, only the
correlations that were not significantly greater than zero (p < .05, one-tailed) are reported.

Fisher’s z’ transformation was used to test hypotheses regarding differences
between the correlations of two groups. Because the sampling distribution of Pearson’s r
is not normally distributed, R. A. Fisher developed a transformation that converts
Pearson’s r to the normally distributed variable z’. The formula for the transformation is:

\[ z' = .5[\ln(1+r) - \ln(1-r)] \]

where \( \ln \) is the natural logarithm. Using this transformation, the equality of two
population correlations can be tested. Here, the null hypothesis H₀: ρ₁ = ρ₂ is tested
against the alternative hypothesis H₁: ρ₁ ≠ ρ₂. The alternative hypothesis H₁: ρ₁ > ρ₂ was
used if the test was one-tailed. This transformation was used to test the three hypotheses
put forth in the introduction.

First, the hypothesis that better readers are more accurate at monitoring their
comprehension than poor readers was examined. Reading ability was operationalized as
the participants’ scores on the Comprehension subtest of the NDRT. Those who scored
above the median were labeled as better readers, and those below the median were
labeled as poor readers. The Comprehension subtest scores ranged from 19 to 38 (out of
a possible 38), and the median was found to be 32. Those whose scores were at the median were randomly placed into one of the two groups.

Metacomprehension accuracy for each group (i.e., better vs. poor readers) was found by computing Pearson’s $r$ correlations using each individual’s median comprehension test percentage correct and median comprehension rating. As seen in Table 3.1, the metacomprehension accuracies for each group were not significantly different from each other ($p = .46$), although the better readers showed a slightly higher correlation than the poor readers.

<table>
<thead>
<tr>
<th>Metacomprehension Accuracy</th>
<th>Better Readers</th>
<th>Poor Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r = .422$</td>
<td></td>
<td>$r = .405$</td>
</tr>
<tr>
<td>$n = 49$</td>
<td></td>
<td>$n = 52$</td>
</tr>
</tbody>
</table>

Table 3.1: Comparison of metacomprehension accuracy (measured by Pearson’s $r$ correlations) for each reading group level.

Next, the second hypothesis was tested. This hypothesis predicted better metacomprehension accuracy for the Delayed Summary condition than for the other two conditions (i.e., Immediate Summary and No Summary conditions). Metacomprehension accuracy was operationalized in the same way as described above. As shown in Table 3.2, the two summary conditions did demonstrate higher metacomprehension accuracies than the No Summary condition. And although the difference between the correlations for the Delayed Summary condition and the No Summary condition approached significance ($p = .08$, one-tailed), no other significant differences were found.
<table>
<thead>
<tr>
<th>Metacomprehension Accuracy</th>
<th>No Summary</th>
<th>Immediate Summary</th>
<th>Delayed Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r = .29$</td>
<td>$r = .562$</td>
<td>$r = .563$</td>
</tr>
<tr>
<td></td>
<td>n = 37</td>
<td>n = 28</td>
<td>n = 36</td>
</tr>
</tbody>
</table>

Table 3.2: Comparison of metacomprehension accuracy (measured by Pearson’s $r$ correlations) for each condition.

Finally, the third hypothesis predicted that delayed summarization would differentially improve metacomprehension accuracy based on reading ability. Metacomprehension accuracy was operationalized in the same way as described above. As shown in Table 3.3, both summary conditions did demonstrate better metacomprehension accuracy than the No Summary condition. The No Summary conditions were, in fact, not significantly different from zero ($ps > .05$ in both better and poor readers), indicating that rating of comprehension did not predict subsequent test performance. There was a significant difference between the No Summary condition and the Delayed summary condition for the poor readers ($p < .05$). Also, the difference between the No Summary condition and the Immediate Summary condition for the better readers did approach significance ($p = .08$, one-tailed). No other significant differences were found.
<table>
<thead>
<tr>
<th>Metacomprehension Accuracy</th>
<th>No Summary</th>
<th>Immediate Summary</th>
<th>Delayed Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better Readers</td>
<td>$r = .400$</td>
<td>$r = .742$</td>
<td>$r = .549$</td>
</tr>
<tr>
<td></td>
<td>$n = 17$</td>
<td>$n = 15$</td>
<td>$n = 18$</td>
</tr>
<tr>
<td>Poor Readers</td>
<td>$r = .182$</td>
<td>$r = .540$</td>
<td>$r = .650$</td>
</tr>
<tr>
<td></td>
<td>$n = 20$</td>
<td>$n = 13$</td>
<td>$n = 18$</td>
</tr>
</tbody>
</table>

Table 3.3: Comparison of metacomprehension accuracy (measured by Pearson’s $r$ correlations) for each reading group level and condition.

As aforementioned, in the present study metacomprehension accuracy was operationalized as the correlation between the median comprehension rating and the median percentage correct on the comprehension tests. Because this measure was used, the individual texts were examined to determine if there were any differences in the metacomprehension accuracies found for each text. Here, each text was measured separately by computing the Pearson’s $r$ correlation coefficient between the percentage correct on the comprehension test and the comprehension rating for each text. Percentage correct was used as a measure for the comprehension tests in order to relate to results from prior research. Correlations are reported in Table 3.4. Significant differences were found between *The Incas* and *Colosseum*, and *Stonehenge* and *Colosseum* ($ps < .05$).
<table>
<thead>
<tr>
<th>Metacomprension Accuracy</th>
<th>The Incas</th>
<th>Colosseum</th>
<th>Stonehenge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r = .516$</td>
<td>$r = .222$</td>
<td>$r = .474$</td>
</tr>
<tr>
<td></td>
<td>$n = 101$</td>
<td>$n = 101$</td>
<td>$n = 101$</td>
</tr>
</tbody>
</table>

Table 3.4: Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each text.

Due to the significantly lower metacomprension accuracy for the *Colosseum* text, an investigation was conducted to determine whether it would be appropriate to include the data from that text in the analysis. Although there were no longer any significant differences, the trends were still in the same direction. Tables 3.5, 3.6, and 3.7 show the metacomprension accuracies demonstrated without the inclusion of the *Colosseum* text. The importance of this manipulation will be talked about in the Discussion section of this paper.

<table>
<thead>
<tr>
<th>Metacomprension Accuracy</th>
<th>Better Readers</th>
<th>Poor Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r = .524$</td>
<td>$r = .513$</td>
</tr>
<tr>
<td></td>
<td>$n = 48$</td>
<td>$n = 52$</td>
</tr>
</tbody>
</table>

Table 3.5: Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each reading group level (without *Colosseum*).
<table>
<thead>
<tr>
<th>Metacomprension Accuracy</th>
<th>No Summary</th>
<th>Immediate Summary</th>
<th>Delayed Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r = .453$</td>
<td>$r = .678$</td>
<td>$r = .561$</td>
</tr>
<tr>
<td></td>
<td>n = 37</td>
<td>n = 28</td>
<td>n = 36</td>
</tr>
</tbody>
</table>

Table 3.6: Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each condition (without Colosseum).

<table>
<thead>
<tr>
<th>Metacomprension Accuracy</th>
<th>No Summary</th>
<th>Immediate Summary</th>
<th>Delayed Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Better Readers</td>
<td>$r = .558$</td>
<td>$r = .666$</td>
<td>$r = .496$</td>
</tr>
<tr>
<td></td>
<td>n = 17</td>
<td>n = 15</td>
<td>n = 18</td>
</tr>
<tr>
<td>Poor Readers</td>
<td>$r = .355$</td>
<td>$r = .697$</td>
<td>$r = .612$</td>
</tr>
<tr>
<td></td>
<td>n = 20</td>
<td>n = 13</td>
<td>n = 18</td>
</tr>
</tbody>
</table>

Table 3.7: Comparison of metacomprension accuracy (measured by Pearson’s $r$ correlations) for each reading group level and condition (without Colosseum).

In order to discover the mechanism behind the increase in metacomprension accuracy due to delayed summarization, the nature of the summaries were examined. The activation theory of text comprehension (Fletcher, van den Broek, & Arthur, 1996; Thiede & Anderson, 2003; van den Broek, Risden, Fletcher, & Thurlow, 1996) can be used to explain why the timing of summary writing might affect metacomprension accuracy. According to this theory, spreading activation in mental networks (i.e., working memory) occurs during cognitive processes like reading. However, this activation is temporary; thus, more information is active in working memory immediately after reading a text than after a delay (during which the activation decays). So, if a
person was to write a summary immediately after reading a text, her mental network
would be highly activated. Hence, she would have access to all of the activated
information in WM when writing a summary, even if the text was not well comprehended
or deeply encoded into LTM. Because of this unlimited access to activated information,
she may base the summary on the extraneous details that were activated during reading.
However, these details may not be accessible when her comprehension is tested due to
the decay of the contents of the WM (Nelson & Dunlosky, 1992). Therefore, because a
person may have access to extraneous details when writing a summary immediately after
reading, the summaries of well-understood texts and less-understood texts may seem
quite similar. In other words, a person may feel a sense of comprehension when details
are easily accessible during summary writing, although it may be unclear whether or not
this sense of comprehension is due to genuine comprehension. Therefore, writing
summaries immediately after reading texts may create misleading cues for judging
comprehension. These cues may, in turn, make it more difficult to distinguish well-
understood texts from less-understood texts. Furthermore, these cues may not be
predictive of future test performance, given that the test occurs after a delay (during
which WM decays). Therefore, one might expect details to be a prominent feature in
summaries that are written immediately after reading. This is in contrast to what one
would expect of summaries that are written after a delay. Here, information stored in
WM may have decayed and so only the information stored in LTM will be accessible
during summary writing. In this case, for a less-understood text, a person may have little
to draw on when writing a summary; whereas, for a well-understood text, more
information may be accessible when writing a summary. Accordingly, writing
summaries after a delay may produce authentic cues for judging comprehension that can highlight the differences between well-understood texts and less-understood texts. Moreover, these cues are likely to be highly predictive of future test performance because both the summary writing and the testing occur after a delay and so are both based on retrieval of information from LTM. And because information stored in LTM is more often in the form of gist than in verbatim details, one might expect more main points or overarching themes (i.e., gist) to be found in summaries that are written after a delay.

As a post hoc test, the features present in the participants’ summaries were examined. Specifically, the amount of main points and details found in summaries written immediately or after a delay were compared. The impact of reading ability on the features of the summaries were also examined. Here, the comparison was between the number of main points and details found in summaries written by better readers (i.e., those whose NDRT Comprehension subtest score was above the mean) and by poor readers (i.e., NDRT Comprehension subtest score below the mean).

In order to quantitatively examine the features of the summaries, a coding system based on sums of main points and details was devised. First, the lists developed by Jelinek-Lewis and Jackson (2001) of the main points, relevant details, and irrelevant details contained in each text were used to tally the number of main points and relevant details mentioned in each summary. Thus, each summary was described by a Main Point score and a Details score. A sample summary and its coding scores are provided in Appendix B.

In order to compare the features of the summaries written immediately to summaries written after a delay, and also to compare summaries written by better and
poor readers, two analyses of variance (ANOVAs) were conducted. Independent variables for both were reading ability (high or low) and condition (Immediate Summary or Delayed Summary) while the dependent measures were number of main points in one ANOVA and number of details in the other. Main effects in the number of details were found for condition and reading ability ($p < .05$). No significant effects were found for the number of main points. Figure 3.1 shows the comparison of condition and reading ability in the number of details contained in the summaries. In sum, participants in the Immediate Summary condition included significantly more details in their summaries than participants in the Delayed Summary condition, and better readers were found to include significantly more details in their summaries than poor readers regardless of condition.

![Graph showing comparison of details](image)

**Figure 3.1:** Comparison of the means (and standard errors) of the number of details for better and poor readers in the Delayed and Immediate Summary conditions.
CHAPTER 4

DISCUSSION

The present study investigated some of the factors that influence the accuracy of monitoring, in terms of reading comprehension. The metacomprehension accuracy paradigm provided us with a means of investigating the development of metacognitive processes in the context of reading. This paradigm examines the relation between learners' judgments of comprehension and their performance on comprehension tests. In other studies, summarization of texts prior to comprehension tests improved learners' metacomprehension accuracy; however, this effect was only shown to be true when a delay was introduced between study and judgments of comprehension. The literature also suggests that better readers demonstrate higher metacomprehension accuracy and are more capable of using metacognitive strategies to their benefit. The present study attempted to replicate these results while also examining the interactions of reading ability with the writing of summaries and introduction of a delay. In addition, the effects that the individual texts may have had on metacomprehension accuracy were also considered. Finally, the features contained in the summaries were analyzed in order to determine a cognitive mechanism behind the effects seen with the introduction of summarization and delays.
4.1: IMMEDIATE AND DELAYED SUMMARIZATION EFFECTS

Although previous research (Thiede & Anderson, 2003) has found a positive effect of delayed summarization on metacomprehension accuracy, the results of the present study indicate only an effect of summarization. That is, participants who summarized the texts before judging their comprehension were more accurate at monitoring their comprehension than participants who did not summarize, but there was no difference in accuracy between those who summarized immediately and those who summarized after a delay. This finding suggests that the key to the delay effect may be in the length of the delay. Many studies using paired associate tasks (Dunlosky & Nelson, 1994; Kennedy & Yorkston, 2000; Nelson & Dunlosky, 1991) came to the conclusion that the minimum delay needed was 30 seconds – long enough to for information in the WM to decay. However, the present results (along with those in Maki, 1998) do not find this to be the case. One reason for this disparity could be due to the length of the texts. The average length of the texts used in this study is much longer than the lengths of the texts used in many other studies of metacomprehension accuracy (Maki & Berry, 1984; Nelson & Dunlosky, 1991; Thiede & Anderson, 2003). Perhaps this could indicate an inverted U-shaped curve in the effect of delay on metacomprehension accuracy. In other words, delays of 30 seconds do show improved accuracy over immediate judgments, but this positive effect may decrease as the delay increases. This could be easily investigated in future studies by designing experiments including several conditions with varying lengths of delay – from immediate through several minutes, hours, or even days.
4.2: EFFECTS OF READING ABILITY

Previous research has revealed a positive effect of reading ability on metacomprehension accuracy; that is, better readers are more accurate at monitoring their own comprehension than poor readers (Glover, 1989; Hacker et al., 2000; Maki & Berry, 1984). However, the present study did not support these results; although better readers had slightly higher metacomprehension accuracy, this difference was not significant. The reason for the nonsignificant results could be due to the nature of the participants in this study. Only university freshmen participated in this study, and this may have reduced the variance of reading ability. This is because university freshmen must have at least a certain reading ability level in order to be admitted to the university, so the present sample did not include individuals who scored very low on standard tests of reading ability.

However, although the variance of reading ability was reduced due to sample constraints, significant interactions were found between reading ability, delay, and summarization. Specifically, poor readers demonstrated higher metacomprehension accuracy when they summarized texts after a delay than when they did not summarize or summarized immediately after reading. On the other hand, better readers were more accurate at monitoring their comprehension when they summarized immediately after reading than when they did not summarize or summarized after a delay. Therefore, the third hypothesis posed earlier in this paper was supported. That is, summarization after a delay improved metacomprehension accuracy for both groups of readers, but there was a greater significant difference in the size of the improvement for the poor readers, as opposed to that of the better readers.
The educational implications of this finding are clear. Poor readers’ metacomprehension accuracy could benefit from the application of an organizational strategy, such as summarization, but only if there is a delay introduced between study and judgment of comprehension. On the other hand, better readers already have these strategies in place, and so such an application may not be beneficial for them. Indeed, the application of such strategies after a delay could even be detrimental to their metacomprehension accuracy, as indicated in the present findings.

4.3: EFFECTS OF TEXT TYPE

In the analysis of effects of individual texts on metacomprehension accuracy, significant differences were found. This was surprising due to the similar nature of the texts. However, the Colosseum text produced very different results. There are several possible reasons for this difference. First, the Colosseum text did have a higher readability score (i.e., Flesch-Kincaid score of 10.0) than the other two texts. This slight increase in readability could have overwhelmed the reading ability levels of the participants in this study, causing their metacomprehension accuracy levels to plummet, compared to the other two texts (for a discussion of the impact of text difficulty on metacomprehension accuracy, see Weaver & Bryant, 1995).

Another possible explanation is the nature of the texts themselves. The texts used in this study were excerpts from the transcripts of narrated documentaries shown on television. And although these three texts seem to be very similar types of texts, there is one critical difference. In the Colosseum documentary (shown on television), the narrator was on-screen while narrating the events that were also on-screen. This is opposed to the other two documentaries where the narrator was not visible during narration; only the
subjects of the narration were present on-screen. This subtle difference could provide an explanation for the difference in metacomprehension accuracy in the three texts. While a narrator is on-screen, he may not include as much description in his narration as a narrator who is off-screen. On-screen narrators may use gestures or other visible cues to add to their descriptions of the subjects of the narration. These visible cues would, of course, not be included in the transcript, and therefore the transcripts could include gaps in the descriptions that could be detrimental to reading comprehension. In other words, it could be that having an on-screen narrator causes the transcripts of the narrations to omit important cues, thereby increasing the difficulty a reader would have in comprehending such a transcript. Future research on this topic could be designed to test these possibilities and would provide a needed addition to the metacomprehension literature – the impacts of various types of texts on metacomprehension accuracy.

4.4: FEATURES OF SUMMARIES

In order to determine the mechanism behind the improvement in metacomprehension accuracy due to summarization, the content of the summaries was analyzed. Specifically, the organizational features of the summaries (i.e., main points and relevant details) were tallied and compared across groups. Although there were no significant differences in the number of main points between groups, there were significant differences in the number of details. Main effects for both summary condition and reading ability were found in the analysis of the details. That is, participants in the Immediate Summary condition included significantly more relevant details in their summaries than participants in the Delayed Summary condition, and better readers
included significantly more details in their summaries than poor readers, regardless of condition. The implications of these main effects will be addressed individually.

First, the main effect of summary condition (i.e., summaries written immediately after reading a text include more details than summaries written after a delay) supports the activation theory of text comprehension (Fletcher, van den Broek, & Arthur, 1996; Thiede & Anderson, 2003; Thiede, Anderson, & Therriault, 2003; van den Broek et al., 1996). Again, this theory states that information in WM activated during reading is still accessible immediately after reading, but decays and is no longer accessible after a delay. Thus, according to this theory, immediate summaries should include more specific details that were activated during reading. Delayed summaries may not include as many specific details because some details may not be encoded into LTM and therefore will have decayed from WM by the time the summary is written.

The main effect of reading ability has a related explanation. Better readers tend to include more details in their summaries because they also tend to have larger and more efficient WM capacities. That is, research has indicated that better readers are able to extract relevant information faster and more efficiently than poor readers (Brown, Armbruster, & Baker, 1986), are able to suppress irrelevant information in WM (Gernsbacher, 1993), and are faster at updating information in WM than poor readers (Palladino, Cornoldi, De Beni, & Pazzaglia, 2001). Thus, better readers are more likely to include relevant details in their summaries and less likely to include irrelevant details. One aspect of the summaries that was not investigated in this study, but could be the aim of future research, is to examine the irrelevant or extraneous details that readers include
in their summaries. This may be the key to understanding the mechanism behind the differences in better and poor readers’ metacomprehension accuracy.

4.5: CONCLUSIONS

Accurate metacognitive monitoring is critical to learning. Specifically, metacognitive monitoring provides a foundation for making decisions about what to study or how long to study the material. However, because these metacognitive skills are unlikely to develop without proper training in metacognitive skills, it is important to understand the mechanisms behind the development of these skills. In understanding the mechanisms, we can not only provide a fuller picture of metacognition to the psychological community, but also impact educational practice by suggesting interventions that will lead to improved metacognitive skills.

In the present study, significant effects of summarization on metacomprehension accuracy were found. That is, learners who summarized texts before making a judgment of comprehension demonstrated better metacomprehension accuracy than those who did not summarize. But that finding does not tell the whole story. By comparing learners of different reading ability levels, differences were detected in the improvement caused by the summarization. These differences were due to variations in the delay between reading and summarization. This study also investigated the mechanisms behind the summarization and reading ability effects by examining the features of the summaries themselves. Here, the key differences lie in the number of relevant details that are included in the summaries. Future investigations should aim to uncover more factors that underlie these metacognitive differences that were discovered in the present study, thereby adding more details to the current picture of metacognitive processes.


APPENDIX A

SAMPLE TEXT AND COMPREHENSION QUESTIONS
The Incas

The Peruvian Andes of South America are among the most rugged mountain chains on earth. Battered by earthquakes, volcanoes, and powerful storms, the Andes are a dynamic land of environmental extremes. Steamy Amazon jungle quickly gives way to jagged 20,000-foot-high peaks. The spine of the Andes separates arid coastal desert from bleak, high-altitude plateau.

Too dry or too vertical form normal living, this land seems an unlikely place to find a great civilization, but 500 years ago, an ambitious Andean people called the Inca were building spectacular cities in the clouds. Their intrepid engineers linked these mountaintop citadels with a phenomenal system of roads and gossamer-like suspension bridges made only of grass. Five hundred years ago, before the Spanish came to the New World; the Inca Empire was the greatest in the Americas, stretching almost the entire length of the Andes.

The Incas were certainly the strangest and most bizarre civilization that the earth has ever seen. They had none of the things that we think of as the prerequisites for a major civilization – no arch, no wheel, no codified mathematics. They couldn’t write. They couldn’t even scratch down an arithmetic problem and yet they could do this amazing engineering.

The Incas’ engineering medium was stone. In their walls and buildings, they showed a mastery over stone that is unrivaled. Without mortar, the Incas created walls of
interlocking blocks that have successfully withstood earthquakes for centuries. But their
interest in stone went beyond the utilitarian; the Incas worshipped rocks.

They carved intricate designs on natural outcrops and poured chicha maize beer
or sacrificial blood down the channels, to honor their mummified ancestors housed in
rock-cut chambers below. Choosing high vantage points at sacred sites, the Incas created
mysterious stone columns dubbed “hitching posts of the sun.” They worshipped the sun
and may have used the shafts for sighting stars. Blending their stonework into the natural
landscape, the Incas carved rocks to mimic the shape of the mountains behind.

Their most impressive and mysterious stonework is found in the walls of their
citadels: giant blocks, some weighing a hundred tons, sit next to each other so precisely
that not even a razor blade can fit between them. Without iron tools, draft animals or the
wheel, how did the Inca builders move and set such large blocks? To answer this
question, we invited several experts with widely different backgrounds to come to Peru.

Professor of architecture, Jean –Pierre Protzen, studies the Incas’ use of stone.
He’s written a book about Inca architecture and has some definite ideas about their
construction methods. Ed Franquemont is both an anthropologist and a building
contractor who lived in a Peruvian village for several years. His particular interest is
how the Inca builders organized their labor force. Philippe Petit is the man who walked a
tightrope between the towers of the World Trade Center. He wants to know how the Inca
builders used grass to make the strong ropes that support their high suspension bridges
and he’s come here to help build one. Vince Lee is an architect and explorer who has
traveled extensively in the Andes looking for lost Inca sites. He has a theory about how
the Inca stonemasons made such precise joints with such giant stones.
A good place to start looking for clues is the citadel overlooking the town of Ollantaytambo. About 500 years ago, a sun temple was under construction inside the fortress. So with all these blocks of stones here, this is clearly a construction site that was abandoned in progress. The question is, where did these stones come from and how did they get here? Well, they came from the quarries on the other side of the river at the base of a mountain.

The team decides to follow the route to the quarry taken by the ancient stone haulers. The hike will take them down a sloping ramp to the valley floor. Along the way, they find massive blocks abandoned by the Inca workers. The villagers call these rocks *piedras cansadas* – weary stones. One legend tells of stones that grew tired, wept blood and refused to move, and so it was there.

The lower quarry is at the bottom of this rockfall; you can see some ramps. The other quarry is at the foot of the cliff. Jean-Pierre – “J.P.” to his friends – leads the team along the remains of a roadway that leads to the quarry, which is actually a rockfall created by rocks eroding from the cliffs above. Here, they find a 70-ton stone that Inca quarry workers had turned into a rectangular block. J. P. believes that all the boulders were first squared off in the quarry. But how did the Incas transport these heavy blocks down the mountain and up to the sun temple on the other side of the valley? Spanish chronicles tell us that the Incas did not possess the wheel or strong draft animals like oxen.

David Canal, a community leader and Inca descendant, believes they hauled the blocks by hand. He’s organized a team of pullers to transport a one-ton rock along the same route taken by the Incas between the quarry and the citadel. For most of its length,
the ramp has a gentle slope but halfway down the mountain, the incline suddenly turns into an almost vertical 800-foot chute to the valley below. With a block more than ten times the size of this one, it must have been extremely difficult for the ancient stone haulers to negotiate this chute. Unlike the Inca block observed on the transport route, this boulder has not been squared off and it tumbles out of control.
The Incas Questions

Circle the letter next to the best answer.

1. Why do the Andes seem like an unlikely place to find a great civilization?
   a. They are between coastal deserts and high-altitude plateaus.
   b. They are too dry and too vertical for normal living.
   c. They had suspension bridges made of grass.
   d. They had not yet been discovered by the Spanish.

2. What made the Inca cities spectacular?
   a. They were built 500 years before the Spanish came.
   b. The mountains were connected by roads and suspension bridges.
   c. The Inca were the most bizarre civilization on earth.
   d. They were built in the steamy Amazon jungle.

3. The most difficult place to build an entire city would be
   a. In the village.
   b. On a plateau.
   c. On the beach.
   d. On a mountain top.

4. Why were the Incas a strange and bizarre civilization?
   a. They were Indians.
   b. They had no prerequisites for a major civilization.
   c. They lived in the mountains.
   d. They worshipped stars.
5. We consider Inca engineering amazing because
   a. They lived in rugged mountains.
   b. They carved rocks.
   c. They had no arch, wheel, or written mathematics.
   d. They had no help from the mountain animals.

6. The wheel is important for civilization because
   a. It makes it easier to transport objects.
   b. It makes communication easier.
   c. It helps make clothes.
   d. It makes it easier to climb mountains.

7. Where did the stones for the sun temple come from?
   a. The Spanish brought them 500 years ago.
   b. Quarries at the base of the mountain.
   c. The Mayas donated them.
   d. They were a gift from the sun-god.

8. How did the Incas transport the blocks down the mountain?
   a. They were pulled by oxen.
   b. They put them in large trucks.
   c. They hauled the blocks by hand.
   d. They used the Spanish to provide horses.

9. What does the legend of the “weary stones” explain?
   a. How stones have human-like qualities.
   b. That stones have mysterious magical powers.
   c. The stubbornness of the abandoned stones.
   d. Why some massive stones were abandoned.
10. What was the Incas’ engineering material?
   a. Iron.
   b. Stone.
   c. Clay.
   d. Glass.

11. What did the Incas build out of stone?
   a. Wheels.
   b. Ovens.
   c. Columns.
   d. Bridges.

12. The walls the Incas built were
   a. Fragile and breakable.
   b. Eroded and falling apart.
   c. Simple and plain.
   d. Sturdy and strong.

13. Why did the Incas carve rocks to mimic the shape of the mountains?
   a. To scare bad spirits.
   b. To please the stone-god.
   c. To honor their mummified ancestors.
   d. To blend into the natural landscape.

14. The stonework in the walls of the citadel is mysterious because
   a. It has precise joints.
   b. It has grass growing out of it.
   c. It weeps blood.
   d. It is shaped like the sun.
15. Where did the Incas bury their dead?
   a. In the river.
   b. In the soft, grassy plain.
   c. In rock chambers.
   d. In the sun temple.

16. How do we know the sun temple construction site was abandoned?
   a. There are blocks of stones left inside the fortress.
   b. The Spanish wrote about it.
   c. There were Inca legends about construction sites.
   d. The Inca builders went on strike.

17. Where was the quarry?
   a. At the base of a meteor shower.
   b. Inside a volcano.
   c. At the foot of a cliff.
   d. In a man-made rock-fall.

18. Why are the Incas a great and amazing civilization?
   a. They lived in the most rugged mountain chain on earth.
   b. They survived the Spanish invasion.
   c. They worshipped stones and the sun.
   d. They build a city in the mountains without a wheel or strong animals.
APPENDIX B

SAMPLE SUMMARY AND CODING
SAMPLE SUMMARY

The Incas were around 500 hundred years ago before the Spanish came to the New World. Also known as Andean people. They were known for their fascinating engineering, but yet they did not have such utensils as tools, oxens, or even simple inventions like the wheel. Their buildings are a masterpiece and withstood many earthquakes and shakes. They did not have cement or mortar only the rocks that they shaped by themselves.

The stones would have designs on them and were used to be worshipped by the Incas. When researched, they discovered that the Incas would transport ton rocks by hand!
SAMPLE CODING

○ 2 Main Points
  ▪ The Incas had engineering despite lack prerequisites for civilization
  ▪ Experts think the Incas moved the stones by hand

○ 5 Details
  ▪ Incas lived 500 years ago
  ▪ Lived in the Andes
  ▪ Did not have mortar
  ▪ Incas worshipped rocks
  ▪ Buildings withstood earthquakes for centuries