The development of domain-specific and domain-general metacognitive monitoring

Brett Douglas Campbell
University of Nevada, Las Vegas
THE DEVELOPMENT OF DOMAIN-SPECIFIC
AND DOMAIN-GENERAL METACOGNITIVE
MONITORING

by

Brett Douglas Campbell
Bachelor of Science
Weber State University
1990

Master of Science
University of Nevada, Las Vegas
2002

A thesis submitted in partial fulfillment of the requirements for the

Doctor of Philosophy Degree in Learning and Technology
Department of Educational Psychology
College of Education

Graduate College
University of Nevada, Las Vegas
May 2007
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
Dissertation Approval
The Graduate College
University of Nevada, Las Vegas

March 27, 2007

The Dissertation prepared by
Brett Douglas Campbell

Entitled
The Development of Domain-Specific and Domain-General Monitoring

is approved in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Learning and Technology

Examination Committee Chair

Dean of the Graduate College
ABSTRACT

The Development of Domain-Specific and Domain-General Monitoring

by

Brett D. Campbell

Dr. Alice J. Corkill, Dissertation Committee Chair
Associate Professor of Educational Psychology
University of Nevada, Las Vegas

Metacognitive monitoring may be a critical element in self-regulated learning. Two types of metacognitive monitoring have been identified: domain-specific and domain-general. Domain-specific metacognitive monitoring occurs when an individual is monitoring content-specific knowledge. Domain-general metacognitive monitoring occurs in situations when content-specific knowledge is not available. Currently no research is available that examines the developmental differences between domain-specific and domain-general metacognitive monitoring in children. This study attempted to address this issue by asking children in first, fourth, and seventh grade to make item-by-item confidence judgments while providing answers in two domain-specific tasks and two domain-general tasks. Two working memory spans tasks were also employed to control for maturational processes. Domain-specific metacognitive monitoring appeared earlier than domain-general metacognitive monitoring. Both domain-
specific and domain-general metacognitive monitoring appear to benefit from experience because older students were more accurate metacognitive monitors and less overconfident than younger students. Maturational processes likely play a less significant role than experience in student improvement at metacognitive monitoring than previously thought.
TABLE OF CONTENTS

ABSTRACT ....................................................................................................................... iii

LIST OF TABLES ........................................................................................................... vii

ACKNOWLEDGEMENTS ............................................................................................. viii

CHAPTER 1  INTRODUCTION ..................................................................................... 1
  The Purpose of the Study ......................................................................................... 5

CHAPTER 2  REVIEW OF RELATED LITERATURE ....................................................... 9
  Monitoring Metacognitive Knowledge ................................................................. 9
  Implementing Metacognitive Knowledge ............................................................ 20
  Monitoring Between Strategies ......................................................................... 27
  Monitoring Declarative Knowledge .................................................................. 36
  Working Memory .................................................................................................. 46
  Summary and Conclusions ................................................................................. 49
  The Present Study ............................................................................................... 53
  Hypotheses ........................................................................................................... 55
  Predictions ............................................................................................................ 59

CHAPTER 3  METHODOLOGY .................................................................................... 61
  Participants ........................................................................................................... 61
  Materials ................................................................................................................ 61
  Procedure ............................................................................................................... 65

CHAPTER 4  RESULTS ............................................................................................... 67
  Performance Scores ............................................................................................... 67
  Metacognitive Monitoring .................................................................................... 68
  Correlations .......................................................................................................... 71
  Principal Component Analysis ........................................................................... 79
  Multivariate Analysis ............................................................................................ 84

CHAPTER 5  DISCUSSION ......................................................................................... 88
  Review of Results .................................................................................................. 90
  Explanation of Results ......................................................................................... 95
  Limitations of the Present Study ......................................................................... 98
  Future Directions ................................................................................................. 100
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Studies examining monitoring of metacognitive knowledge</td>
<td>11</td>
</tr>
<tr>
<td>Table 2</td>
<td>Studies examining monitoring of strategy implementation</td>
<td>21</td>
</tr>
<tr>
<td>Table 3</td>
<td>Studies examining monitoring between two strategies</td>
<td>30</td>
</tr>
<tr>
<td>Table 4</td>
<td>Studies examining errors in metacognitive monitoring</td>
<td>34</td>
</tr>
<tr>
<td>Table 5</td>
<td>Studies examining monitoring of declarative knowledge</td>
<td>37</td>
</tr>
<tr>
<td>Table 6</td>
<td>Means and standard deviations of percentile rankings</td>
<td>68</td>
</tr>
<tr>
<td>Table 7</td>
<td>Means and standard deviations of CAQ scores</td>
<td>69</td>
</tr>
<tr>
<td>Table 8</td>
<td>Means and standard deviations of gamma coefficients</td>
<td>70</td>
</tr>
<tr>
<td>Table 9</td>
<td>Means and standard deviations of bias scores</td>
<td>71</td>
</tr>
<tr>
<td>Table 10</td>
<td>Correlations of percentile rankings for first grade students</td>
<td>72</td>
</tr>
<tr>
<td>Table 11</td>
<td>Correlations of percentile rankings for fourth grade students</td>
<td>72</td>
</tr>
<tr>
<td>Table 12</td>
<td>Correlations of percentile rankings for seventh grade students</td>
<td>72</td>
</tr>
<tr>
<td>Table 13</td>
<td>Correlations of CAQ scores for first grade students</td>
<td>73</td>
</tr>
<tr>
<td>Table 14</td>
<td>Correlations of CAQ scores for fourth grade students</td>
<td>73</td>
</tr>
<tr>
<td>Table 15</td>
<td>Correlations of CAQ scores for seventh grade students</td>
<td>73</td>
</tr>
<tr>
<td>Table 16</td>
<td>Correlations of gamma coefficients for first grade students</td>
<td>74</td>
</tr>
<tr>
<td>Table 17</td>
<td>Correlations of gamma coefficients for fourth grade students</td>
<td>74</td>
</tr>
<tr>
<td>Table 18</td>
<td>Correlations of gamma coefficients for seventh grade students</td>
<td>74</td>
</tr>
<tr>
<td>Table 19</td>
<td>Correlations of bias scores for first grade students</td>
<td>75</td>
</tr>
<tr>
<td>Table 20</td>
<td>Correlations of bias scores for fourth grade students</td>
<td>75</td>
</tr>
<tr>
<td>Table 21</td>
<td>Correlations of bias scores for seventh grade students</td>
<td>75</td>
</tr>
<tr>
<td>Table 22</td>
<td>One component model for percentile rankings</td>
<td>81</td>
</tr>
<tr>
<td>Table 23</td>
<td>Two component model for percentile rankings</td>
<td>81</td>
</tr>
<tr>
<td>Table 24</td>
<td>Two component model for CAQ scores</td>
<td>82</td>
</tr>
<tr>
<td>Table 25</td>
<td>Two component model for gamma coefficients</td>
<td>82</td>
</tr>
<tr>
<td>Table 26</td>
<td>One component model for bias scores</td>
<td>83</td>
</tr>
<tr>
<td>Table 27</td>
<td>Two component model for bias scores</td>
<td>83</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENT PAGE

I would like to acknowledge and thank the dissertation chair, Dr. Alice Corkill, for her advice and patience. I would also like to thank the other committee members for their individualized assistance, Dr. Gregory Schraw, Dr. CarolAnne Kardash, and Dr. Mark Guadagnoli. I would like to acknowledge my wife, Kay Campbell for her support.
CHAPTER 1

THE DEVELOPMENT OF DOMAIN-SPECIFIC AND
DOMAIN-GENERAL METACOGNITIVE MONITORING

Metacognitive monitoring is a critical element in learning (Brown, Palinscar & Armbruster, 1984; Nelson & Narens, 1990; Lucia, 1994), recall (Schraw, Dunkle, Bendixen, & Roedel, 1995), problem solving (Artzt & Armour-Thomas, 1992; Carr & Jessup, 1995), and self-regulated learning (Butler & Winne, 1995; Winne & Jamieson-Noel, 2002). Metacognitive monitoring is a process that helps a learner analyze on-going progress toward a given goal. For example, an individual may search for relevant background knowledge or experiences, review whether a current strategy is effective, or determine if new strategies are needed. Researchers have focused on four different ways in which metacognitive monitoring may be measured. The first method is to have participants verbalize awareness of monitoring either by responding to monitoring-oriented statements on surveys or through think-aloud sessions. The second method is to teach learners several types of cognitive strategies and to then observe which strategy, if any, is selected and used. The third method is to identify and classify types of errors learners make or to catalog failures and pitfalls learners experience while attempting to achieve a specific task. The final method is to have participants
rate how confident they are that they have completed a task successfully. For example, learners may be asked to rate how well they believe they have learned a piece of information or how confident they are that they will be able to recall a specific fact. The resulting confidence ratings are then compared to actual performance.

Metacognitive monitoring has been documented in children as young as four years old (Cultice, Somerville, & Wellman, 1982; English, 1992; Schneider, 1998). Cultice and colleagues (Cultice, Somerville, & Wellman, 1982), for example, had children estimate the likelihood that they would recognize a face when presented with a photograph. Four year-old children demonstrated elementary metacognitive monitoring on this task. English (1992), as another example, had students varying in age from four to nine solve a series of problems and tracked their metacognitive monitoring. She discovered that as students age, their monitoring patterns become more thorough and complex. The development of metacognitive monitoring, therefore, may be dependent on experience (Bisanz, Vesonder, & Noss, 1978; Brown, & Smiley, 1978; Kuhn, 2000a). Other factors that appear to affect metacognitive monitoring that have been extensively researched include: metacognitive awareness, or the knowledge of one’s own cognitive processes (Corkill & Koshida, 1993; Schraw, Dunkle, & Bendixen, 1995; Schraw & Nietfield, 1998); acquired strategies and skills (Kreutzer, Leonard, & Flavell, 1975; Mevarech, 1995; Myers & Paris, 1978; O'Sullivan & Pressley, 1984); task difficulty (Baker & Anderson, 1982; Campbell
As a result of this research, two different types of metacognitive monitoring have been identified: domain-specific and domain-general. Domain-specific metacognitive monitoring involves the regulation of specific content knowledge, such as math, history, or geography, and related strategies (Schraw, Dunkle, Bendixen, & Roedel, 1995). Domain-specific metacognitive monitoring has been recorded in students working in reading, science, and math (Baker & Brown, 1984; Lucia, 1994; Carr & Jessup, 1995). In particular, domain-specific metacognitive monitoring has been demonstrated to be influential in reading comprehension (Palinscar & Brown, 1984; Jacobs & Paris, 1987; Juliebo, Malicky, & Norman, 1998), the understanding of scientific principals (Lucia, 1994), and the ability to solve math problems (Maqsud, 1998; Mevarech & Kramarski, 2003). The underlying assumption is that domain-specific metacognitive monitoring primarily occurs within specific domains or content areas (Schraw, Dunkle, Bendixen, & Roedel, 1995).

In direct contrast, researchers who study domain-general metacognitive monitoring hold that all metacognitive monitoring falls under one general, all-encompassing metacognitive process. It has been suggested that domain-
general metacognitive monitoring may be involved in determining the degree of
familiarity of a particular domain—such as music history—the selection of
appropriate strategies, and the allocation of cognitive resources. Domain-general
metacognitive monitoring has been considered to be of special import when a
learner is called upon to monitor while working on novel tasks where specific
domain knowledge or skills are likely unavailable (Schneider, Korkel, & Weinert,

Several researchers (e.g., Bisanz, Vesonder, & Voss, 1978; Markman,
1979; McGivern, Levin, Pressley, & Ghatala, 1990; English, 1992; Short,
Schatschneider, & Friebert, 1993; Bartsch & Estes, 1996; Fletcher-Flinn &
Snelson, 1997; Kuhn 2000b; Dunlosky, 2002) have attempted to identify a
developmental trend in metacognition and metacognitive monitoring. At present,
developmental trends exclusive to either domain-specific or domain-general
metacognitive monitoring have not been investigated, though researchers have
suggested that such developmental differences do exist (Schraw & Nietfeld,
1998). The purpose of the present study, therefore, is to take a first step in
attempting to identify developmental differences in domain-specific and domain-
general metacognitive monitoring assuming that differences exist. Such findings
could contribute to a theoretical framework for the understanding of the
development of domain-specific and domain-general metacognitive monitoring in
children. In addition, such findings may aid in the development of metacognitive
strategy instruction. Guidance about when (age-wise) to implement strategy
instruction and which strategies should be taught, domain-specific, domain-general, or both, is sorely needed.

*The Purpose of the Study*

The purpose of this study is twofold. The first goal is to make an attempt to identify the developmental trend of domain-general and domain-specific metacognitive monitoring in children. The second goal is to consider whether domain-specific and domain-general metacognitive monitoring differ during development. With respect to the first goal, two possible developmental trends with empirically derived explanations have been identified: experiential and modular. Experiential theorists would suggest that domain-specific metacognitive monitoring is predominantly a function of experience and that domain-general metacognitive monitoring is a function of strategy transfer across domains. Modularization theorists would suggest that domain-general monitoring is a default process and domain-specific monitoring develops as a result of proficiency within a domain.

The experiential hypothesis is that as students become more proficient in specific content areas strategy knowledge and skills will transfer from specific domains to all domains (Schraw & Nietfeld, 1998). If this is true, domain-specific metacognitive monitoring should appear in fairly young students and domain-general metacognitive monitoring should not appear until later in childhood. In addition, if the experiential hypothesis is true, domain-specific metacognitive
monitoring may be more efficacious, as measured by higher monitoring accuracy scores, than domain-general metacognitive monitoring because learners will have more domain-specific metacognitive monitoring practice.

The modularization hypothesis is that metacognitive monitoring should begin as a domain-general process that becomes domain-specific with experience. This hypothesis stems from neural network models (Karmiloff-Smith, 1992; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Karmiloff-Smith, Plunkett, Johnson, Elman, & Bates, 1998). The modularization hypothesis is based primarily on the language acquisition model of Karmiloff-Smith and Bates (e.g., 1992; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996). This model would support the notion that although brain development has general predispositions, it becomes specialized, or modularized, with exposure to a variety of experiences. If this were the case, domain-specific metacognitive monitoring would develop not only after domain-general metacognitive monitoring but as a result of it.

The second goal of this study is to examine whether domain-specific and domain-general metacognitive monitoring differ during development. Learners show improvement in domain-specific metacognitive monitoring as a result of explicit instruction and practice (Brown, Palinscar, & Armbruster, 1984; Paris & Jacobs, 1984; Lucia, 1994; Mevarech & Kramarski, 2003). In addition, domain-specific metacognitive monitoring may be more accurate than domain-general metacognitive monitoring. It is assumed that instruction and practice is more
likely to occur with domain-specific tasks than with domain-general tasks. Furthermore, without specific instruction and practice domain-general monitoring may be limited to the influence of working memory capacity and the demands of the task (Swanson, 1999).

In order to address the goals of this study, students from first, fourth, and seventh grade were given tasks hypothesized to be domain-specific (arithmetic and reading) as well as tasks hypothesized to be domain-general (the Wisconsin Card Sorting Test and the Raven's Standard Progressive Matrices). Study participants were required to provide confidence ratings for each item in the domain-specific and domain-general tasks as a measure of metacognitive monitoring. Measures of working memory were also collected in order to control for potential developmental differences due to maturation.

Support for the experiential hypothesis would be demonstrated by linear improvement in both domain-specific and domain-general metacognitive monitoring across the three grades (see Figure 1). Greater accuracy in domain-general metacognitive monitoring (relatively speaking) should appear later. Older students should show greater metacognitive monitoring accuracy for both domain-specific and domain-general tasks. Support for the modularization hypothesis would be demonstrated by the presence of both domain-general and domain-specific metacognitive monitoring in first grade participants; domain-specific metacognitive monitoring, however, would be expected to increase sharply across the three grade levels likely as a function of practice (see Figure
2). Students from all three grades should show equivalent metacognitive monitoring accuracy in the domain-general tasks when controlling for maturation.
CHAPTER 2

REVIEW OF RELATED LITERATURE

Four questions have been asked by researchers in an effort to understand metacognitive monitoring: 1) what metacognitive knowledge does a student have; 2) how well is metacognitive knowledge applied; 3) what do you do when you have two strategies to choose from; and 4) how well does an individual monitor his or her own declarative knowledge? A variety of approaches have been used to answer each question: interviews, think-alouds, participant choice, participant testing, and confidence ratings. In this review, each issue will be examined individually.

Monitoring Metacognitive Knowledge

Metacognitive knowledge has been defined as a three component variable which includes: 1) the information a person has about the elements of a task; 2) what the individual knows about his/her characteristics as a learner; and 3) the strategy knowledge that is available to that individual (Flavell & Wellman, 1977; Flavell, 1979). A parallel definition treats metacognitive knowledge as the conditional knowledge available to the person relative to the task (Schraw &
Moshman, 1995). Access and application of this knowledge can facilitate success in completing a task via metacognitive monitoring. Initial attempts at learning about and measuring metacognitive knowledge was through interviews with students. More recently, researchers have observed students' self-talk during learning or problem-solving tasks. The next several sections of this review will consider each of these approaches in turn. The articles described in this section are summarized in Table 1.

Student Interviews

In one of the first metacognition studies, kindergarten, first, third, and fifth grade students were interviewed about their personal awareness of metacognition (Kreutzer, Leonard, & Flavell, 1975). Interview questions covered areas such as individual differences (e.g., "Can you remember better than your friends?"), forgetting (e.g., "Do you forget?"), relearning (e.g., "Jim learned the names of birds, but forgot them. Bill never learned the names of birds. Which one will learn the names of birds faster?"), and memory strategies (e.g., "Will a story help you remember a list of words?"). Kreutzer et al. (1975) concluded that kindergarten and first grade students had a basic understanding of metacognition. In addition, kindergarten and first graders recognized that: 1) there is rapid decay in short-term memory, 2) previously learned information can be forgotten, and 3) retrieval is a function of the amount of study time. First grade students were also able to identify external mnemonic devices (e.g.,
Table 1. Studies examining metacognitive monitoring using interviews and self-talk methodologies.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Grade</th>
<th>Condition</th>
<th>Results</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kreutzer, Leonard, &amp; Flavell</td>
<td>1975</td>
<td>K, 1, 3, 5</td>
<td>Identify metacognitive knowledge</td>
<td>Metacognitive knowledge identified in kindergarten students. Third and fifth grades students recognize individual differences.</td>
<td>Younger children have some metacognitive awareness. Older students make distinctions between person and task.</td>
</tr>
<tr>
<td>Myers &amp; Paris</td>
<td>1978</td>
<td>2, 6</td>
<td>Identify reading strategy knowledge</td>
<td>Second grade students are unaware of strategy knowledge, but sixth grade students have strategy knowledge.</td>
<td>Younger students not sensitive to limitations of tasks or strategies.</td>
</tr>
<tr>
<td>Clift, Ghatala, Naus, &amp; Pool</td>
<td>1990</td>
<td>K-12 Teachers</td>
<td>Identify metacognitive monitoring instruction.</td>
<td>Strategy instruction occurs, but ineffectively.</td>
<td>Poor monitoring may be due to ineffective instruction.</td>
</tr>
<tr>
<td>Short, Schatschneider, Friebert</td>
<td>1993</td>
<td>2, 4, 6</td>
<td>Questioning during cognitive tasks</td>
<td>Steady growth in domain-specific monitoring.</td>
<td>Domain-specific monitoring more influenced by experience than domain-general monitoring.</td>
</tr>
<tr>
<td>O'Sullivan &amp; Joy</td>
<td>1994</td>
<td>1, 3, 5, 7</td>
<td>Report on fictitious students' reading errors</td>
<td>Most students accurately identified problems. Suggested apply more effort as solution. No differences between grades.</td>
<td>Students unable to provide effective remediation solutions.</td>
</tr>
<tr>
<td>Malicky, Juliebo, Norman, &amp; Pool</td>
<td>1997</td>
<td>1</td>
<td>Videotaped student comments during reading instruction</td>
<td>Comments were categorized into self-corrections, familiarity, and ease of reading.</td>
<td>Students have a rudimentary understanding of reading strategies.</td>
</tr>
<tr>
<td>Vanleuvan &amp; Wang</td>
<td>1997</td>
<td>1, 2</td>
<td>Self-monitoring comments during reading and math instruction</td>
<td>Students have few comments related to monitoring. Low achieving students have even less monitoring-related comments.</td>
<td>First and second grade students do not spontaneously engage in monitoring.</td>
</tr>
<tr>
<td>Juliebo, Malicky, &amp; Norman</td>
<td>1998</td>
<td>1</td>
<td>Student comments comparing themselves to</td>
<td>Student monitoring closely resembled model.</td>
<td>Strategy instruction in the form of modeling and tutoring is effective in</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Grade</td>
<td>Condition</td>
<td>Results</td>
<td>Contribution</td>
</tr>
<tr>
<td>--------------------</td>
<td>------</td>
<td>---------</td>
<td>--------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Schneider</td>
<td>1998</td>
<td>Pre-school</td>
<td>Children comments related to recall of lists of words</td>
<td>Students made accurate predictions with familiar words, but overconfident with unfamiliar words.</td>
<td>Overconfidence may represent wishful thinking.</td>
</tr>
<tr>
<td>Veenman, Elshout, &amp; Meijer</td>
<td>1997</td>
<td>College</td>
<td>Judges rated monitoring on students comments in 3 content areas</td>
<td>Identified domain-general and domain-specific monitoring.</td>
<td>Domain-general monitoring has a separate contribution to learning from ability.</td>
</tr>
<tr>
<td>Veenman &amp; Verheij</td>
<td>2003</td>
<td>College</td>
<td>Judges rated monitoring on students comments in 2 content areas</td>
<td>Domain-general factor identified.</td>
<td>Greater support for domain-general monitoring, than for domain-specific monitoring.</td>
</tr>
</tbody>
</table>
records) and recognized that other people may assist with retrieval. Third and fifth grade students could identify differences between individuals in terms of ability as well as differences across and between tasks. Furthermore, both fifth and third grade students recognized the utility of study strategies such as intentional clustering of items and rehearsal. Participants from all four grades identified rehearsal as the primary strategy for remembering.

Only one other study during the early years of metacognitive research considered participant reports on metacognitive awareness. Myers and Paris (1978) interviewed second and sixth grade students and asked them to discuss what they were aware of when they read. Second grade students reported that information is more easily recalled when the text is shorter and that familiar texts are easier to remember than new texts. Second grade students did not express awareness of semantic features, such as organization, goals of reading, or cognitive strategies that assist comprehension. Sixth grade students reported the same task variables identified by second graders. In addition, they were able to discuss issues related to personal ability and cognitive strategies as influences on reading comprehension.

More recently, Mevarech (1995) suggested that kindergarten students with greater metacognitive awareness in mathematics tend to do better with mathematical word problems. In his study, kindergarten students were interviewed about metacognitive awareness and tested on mathematical word problems. General ability, based on teacher judgment, was controlled.
Kindergarten participants gave some indication of metacognitive awareness reporting things such as bigger numbers are more difficult to work with than smaller numbers. Higher levels of metacognitive awareness correlated positively with better performance in mathematics. Furthermore, metacognitive knowledge explained more variance in math performance than did general ability.

One set of researchers looked at age and skill differences in memory and metacognition of students in the second, fourth, and sixth grades (Short, Schatschneider, & Friebert, 1993). In addition to separating students by grade, students were separated by ability (low and average) via digit span and word list recall tasks. Students were given various tests including word knowledge, the Stroop Color/Word test, a matrix memory test (recalling letters within a matrix), and two digit span tests. Differences between task-specific (i.e., domain-specific) and domain-general metacognitive monitoring were also examined. Interviewing students about the matrix memory task and the two digit span tasks was used to measure task-specific metacognitive monitoring. Domain-general metacognitive monitoring was measured using a test of metamemory strategy and taxonomic knowledge (Schneider, Borkowski, Kurtz, & Kerwin, 1986). Students in the average achievement group showed steady growth in both strategic and taxonomic metacognitive awareness across the three grades. Students in the low achieving group did not show any improvement in metacognitive knowledge. Specifically, the low achieving students recalled less information, did not spontaneously use strategies, underutilized organizational strategies, and
acquired task-specific knowledge more slowly than the average students.

Additional analyses indicated that task-specific metacognition was the best predictor on number of words recalled, as well as performance on the matrix memory and digit span tasks. Domain-general metacognitive monitoring was the next best predictor on the memory tasks. Age and ability measures had weak predictive power. The authors concluded that domain-general metacognitive monitoring might be less influenced by experience than domain-specific metacognitive monitoring.

In a study with a slightly different perspective, students were asked to help a fictitious student with schoolwork. O'Sullivan and Joy (1994) interviewed first, third, fifth, and seventh grade students. Four fictitious students with reading comprehension problems were described to each participant. The research participant was asked to determine the cause of the reading problem and provide suggestions for each fictitious student. Most participants identified the reading problem accurately, which O’Sullivan and Joy interpreted as a demonstration of metacognitive awareness. Participant suggestions for remediation, however, were not sophisticated. The predominant suggestion, regardless of the age of the participant supplying the suggestion, was that the fictitious student should apply more effort.

Clift and colleagues (Clift, Ghatala, Naus & Pool, 1990) surveyed elementary and secondary teachers with respect to their perceptions of strategy instruction and how this might be related to metacognitive knowledge in
children. The teachers reported that they engaged in specific learning strategy instruction; however, the strategies were not always taught effectively. The most common learning strategy reported being taught was repetition. Teachers surveyed did request greater exposure to a wider realm of strategy instruction and general metacognitive knowledge including monitoring instruction.

Self-talk

Recently, students' comments to themselves have been examined. Vanleauvan and Wang (1997) recorded elementary school students during reading and mathematics instruction in the classroom. Student spontaneous private speech, or self-interrogations, of in-class tasks were taped and categorized. From the total number of self-interrogations (n = 56), 28% were related to self-monitoring. Students' self-interrogations were categorized further into categories indicative of whether the material had been learned previously or if errors in comprehension had occurred. Boys made more comments related to self-monitoring than did girls. Low achieving students made fewer self-interrogation comments. Although it was not statistically significant difference, Vanleauvan and Wang report that there were more self-interrogations in math than in reading.

In a more detailed examination of students' metacognitive knowledge of task constraints, Malicky, Juliebo, Norman, and Pool (1997) recorded and categorized first grade student's comments during reading instruction. Most students' comments were related to self-corrections. Other metacognition
comments related to the participant's familiarity with the book or how difficult the book was to read. Malicky et al. report that the first graders made more metacognitive remarks when the material was familiar and when the book was easier to read.

In an extension of this study, Juliebo, Malicky, and Norman (1998) showed first grade students models who implemented specific reading strategies in one-on-one tutoring sessions. The students were then filmed reading and were compared to the model. The researchers discovered that the strategies used by the first grade students matched those of the models fairly closely. Specifically, the first graders engaged in more self-corrections and made more comments related to procedural and strategic awareness after tutoring. The authors suggested that even in the early stages of reading, learners could benefit from some form of strategy instruction.

In an investigation of the emergence of metacognition in early childhood Schneider (1998) analyzed the private comments of pre-schoolers. After listening to a list of words, students were instructed to tell the researcher to turn the tape off when they thought they had heard all of the words they would be able to remember. The students were also asked to make predictions as to how well they would recall familiar or unfamiliar words. Students made more accurate predictions and had better recall with familiar words. The preschoolers showed higher levels of overconfidence with unfamiliar words. Schneider interpreted this overconfidence as wishful thinking. Students would tell the researcher they
would do well; in their recorded private speech, however, the children made comments indicating the opposite.

Veenman, Elshout, and Meijer (1997) distinguished between domain-general and domain-specific metacognitive monitoring through student think alouds. Domain-general metacognitive monitoring was referred to as monitoring in tasks without specific content knowledge, while domain-specific metacognitive monitoring referred to monitoring of knowledge in a content area where specific knowledge had been previously learned. In addition, Veenman and colleagues looked at the influence of intellectual ability on monitoring. They gave fourteen college students three problems from three different domains to solve: physics, statistics, and a fictitious domain, calometry. Students did not have sufficient background knowledge for either physics or statistics and calometry represented a new domain for all students. Metacognitive monitoring was measured by analyzing student comments while solving problems. Principal component analysis defined four components: a domain-general monitoring component and three domain-specific components, statistics, physics, and calometry. Veenman and colleagues suggested that intellectual ability had a weak effect on metacognitive monitoring. Based on the analysis of this data, the authors suggested that intellectual ability and domain-general monitoring played independent roles in learning especially when attempting to acquire knowledge or skills in a new domain.
Veenman and Verheij (2003) extended Veenman, Elshout, and Meijer's (1997) study, but utilized only two domains, mathematics and the fictitious domain, calometry. In this instance, mathematics represented a familiar domain while calometry represented an unfamiliar domain. Differences in performance on tests in each domain were compared to intellectual ability and self-reported grade point average. Judges provided ratings on students for both tasks based on student's verbalizations while solving the tasks. The think aloud was used to measure metacognitive monitoring skillfulness. The single component that emerged from a principal component analysis was interpreted as evidence for domain-general metacognitive monitoring. Metacognitive regulation had a low correlation with intellectual ability. Strong, positive correlations occurred between metacognitive measures and performance measures on the two tasks. Furthermore, metacognitive monitoring accounted for more variance than intellectual ability.

All of these studies focused on what could be learned about metacognition by listening, in one form or another, to what research participants said about their metacognitive knowledge and/or processes. Taken together, these studies provide evidence that: 1) metacognitive monitoring and knowledge likely begins as early as pre-school and kindergarten, 2) older students are more proficient monitors than younger students likely due to greater experience with monitoring as well as inadequate monitoring training in the curriculum in the earlier grades,
3) specific experiences facilitates domain-specific monitoring, but not domain-general monitoring, and 4) instruction can facilitate metacognitive monitoring.

Implementing Metacognitive Knowledge

Monitoring a learning strategy occurs in a number of stages during learning. A student may monitor a new learning strategy for effectiveness or in order to generate feedback to improve that strategy's utility. A student may access what they have stored related to a learning strategy that has been monitored in order to select the most appropriate learning strategy for a particular task. A student may monitor a learning strategy in a new situation. Finally, students may monitor when they fail at a task, such as comprehending during reading, finding the main point, or being unable to solve a particular problem. Research on metacognitive monitoring and the implementation of metacognitive knowledge may be categorized into three groups: 1) use of specific metacognitive knowledge, 2) selecting between two strategies, and 3) detecting errors. Each of these areas will be discussed in turn.

Specific Metacognitive Knowledge Use

Successful implementation of metacognitive knowledge can facilitate learning. Proper deployment and use of metacognitive knowledge may speed information acquisition or enhance recall, thus freeing cognitive resources. Appropriate use of metacognitive knowledge may facilitate self-regulation. Research related to the implementation of metacognitive knowledge typically consists of observing
Table 2. Studies examining metacognitive monitoring of strategy implementation.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Grade</th>
<th>Conditions</th>
<th>Results</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masur, McIntyre, &amp;</td>
<td>1973</td>
<td>1, 3, College</td>
<td>Select list of words for study</td>
<td>First grade students selected lists at random. Third grade and college students studied words they did not know.</td>
<td>Improvement in monitoring can be associated to experience to a degree.</td>
</tr>
<tr>
<td>Flavell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brown &amp; Smiley</td>
<td>1978</td>
<td>5, 7-8, 11-12,</td>
<td>Strategy use in finding main idea</td>
<td>Fifth grade students did not utilize strategies as effectively as other grades.</td>
<td>Fifth grade students lack metacognitive knowledge and were less effective in monitoring.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>College</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Sullivan &amp; Pressley</td>
<td>1984</td>
<td>5, College</td>
<td>Type of instruction to facilitate transfer</td>
<td>Elaborated instructions facilitated transfer better.</td>
<td>Explicit instructions and reasoning facilitates strategy transfer.</td>
</tr>
<tr>
<td>Artzt &amp; Armour-Thomas</td>
<td>1992</td>
<td>7</td>
<td>Solve math problems in a group setting</td>
<td>Students with higher metacognitive awareness lead the group to the answer.</td>
<td>Students with greater metacognitive awareness promote greater understanding for the entire group.</td>
</tr>
<tr>
<td>English</td>
<td>1992</td>
<td>Pre-school, K, 1, 2,</td>
<td>Dress bears in all possible combinations</td>
<td>Strategies became more complex and effective with age.</td>
<td>Metacognitive strategy increases in sophistication with age</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3, 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuhn</td>
<td>1995</td>
<td>4, college</td>
<td>Strategy transfer across two content areas</td>
<td>Both grades transferred the strategy with college students adapting quicker.</td>
<td>Strategy transfer occurs with practice in both content areas.</td>
</tr>
<tr>
<td>Carr &amp; Jessup</td>
<td>1997</td>
<td>1</td>
<td>Strategy adaptation in math across a school year</td>
<td>Strategy selection move from external cues to internal strategies</td>
<td>Math strategies are internalized with familiarity.</td>
</tr>
<tr>
<td>Fletcher-Flinn &amp;</td>
<td>1997</td>
<td>Pre-school</td>
<td>Relationship between metacognitive knowledge and</td>
<td>Metacognitive knowledge contributed the most to reading achievement.</td>
<td>Metacognitive awareness is a prerequisite for reading and also a consequence of reading. Reading is part of a domain-general ability</td>
</tr>
<tr>
<td>Snelson</td>
<td></td>
<td></td>
<td>reading achievement</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
student reactions to various problems or by observing the transfer of strategy knowledge from one content area to another. The research described in this section is summarized in Table 2.

In one of the first studies on metacognition, first grade, third grade, and college students were given lists of words to memorize, were tested, and then were allowed to select half of the list to study for a second test (Masur, McIntyre, & Flavell, 1973). First grade students appeared to make their selections at random. The third grade and college students, however, selected words they did not recall the first time.

Brown and Smiley (1978) compared differences in monitoring comprehension strategies between fifth graders, seventh and eighth graders, eleventh and twelfth graders, and college students. Each student read two Japanese folk tales. They were instructed to either underline important aspects of the tale, take notes, or were given no instructions. The fifth, seventh and eighth grade students did not focus solely on important aspects of the tales when using reading strategies such as underlining or note taking. These students were able to identify the main ideas by underlining or note taking, but they underlined or noted trivial aspects of the passage as well. Most fifth graders did not use the underlining strategy effectively. As expected, the older students recalled more information than the younger students. The authors concluded that the younger students were less introspective, less conscious about the
workings of the mind, and less able to exert control over their cognitive processes.

English (1992) created a research scenario that allows a glimpse of three developmentally different strategies used by children to achieve a simple task. Research participants were children between the ages of four and nine. The participants were given six teddy bears and an assortment of brightly colored shirts and pants. The participants were asked to dress the bears in all possible color combinations. The youngest participants (4 and 5 year old children) used a "non-planning" strategy. These participants appeared to dress the bears randomly with no attempt to track color combinations of shirts and pants. Participants between the ages of 5 and 7 used what English termed a "transitional strategy." These children attempted to recall, with varying degrees of success, which color combinations had been used. Participants between the ages of 7 and 9 used an "odometer strategy." In this approach, children would use the same item (red top) until all combinations had been exhausted and then move on to a new combination (red top/blue pants, red top/yellow pants, red top/green pants, blue top/blue pants). The strategies used from ages 4 and 5, to ages 5 to 7, and to ages 7 to 9 clearly increase in complexity and plainly illustrate the differences in monitoring ability between the three age groups.

Artzt and Armour-Thomas (1992) "listened in" as seventh grade students worked in groups while solving math problems via discussion. Students were separated into two categories of metacognitive awareness, high or low, based on
their contributions to the group discussions. Students with higher metacognitive awareness tended to lead the conversations within the groups. Artzt and Armour-Thomas suggested students with greater metacognitive awareness promoted understanding for the group and lead the group in finding the solution for the word problem.

One study has considered whether gender differences would be evident in use of metacognitive strategies (Carr & Jessup, 1997). First grade students were given addition and subtraction problems to solve and were then questioned about the strategies they used. The students' use of strategies was followed over the course of the school year. Girls were more likely to use counters or fingers; boys were more likely to use retrieval to solve the problems. When the problems were presented in social settings, students were less likely to rely on counters. This was interpreted to suggest that metacognitive knowledge and social settings influenced strategy selection. Students' justifications for their strategy selections included availability, usefulness, and strategy capacity. Students with higher awareness relied on retrieval. Metacognitive knowledge increased overall between October and May. Retrieval became more prevalent as the strategy of choice by May. There was no significant difference in metacognitive knowledge between boys and girls.

Fletcher-Flinn and Snelson (1997) examined metacognitive and academic abilities in preschool children. Four-year old children were assessed on metalinguisitic ability (familiarity with the names of objects), linguistic ability
(word recognition), general aptitude (block design and vocabulary recognition from an intelligence test), and social metacognition via a false belief task. The false belief task consisted of a researcher hiding an object in front of the child and a research accomplice. After the accomplice left, the researcher moved the object to another hiding place. The child was asked where the accomplice would look first for the hidden item. The metalinguistic task and the social metacognition tests correlated positively with each other. Fletcher-Flinn and Snelson interpreted the results an indicator that children as young as four may demonstrate general metacognitive awareness. One year later, reading achievement tests were administered. Reading achievement was positively correlated with the metacognition tasks (the metalinguistic task and the social metacognition tests and general aptitude). The authors concluded that the development of metacognition follows a domain-general route.

In his review of the literature, Flavell (1979) recognized that young children have basic metacognitive knowledge; however, young children are not able to use this knowledge to make cognitive tasks easier. While older students tend to do better at metacognitive monitoring than the preschool and kindergarten students, Flavell suggested that experience alone might not be enough to explain this improvement in metacognitive monitoring. One reaction to Flavell’s suggestion was to attempt to teach students to transfer strategies from one content area to another content area.

*Strategy Transfer*
Kuhn (1995) examined metacognitive strategy transfer across two content domains. Fourth grade and community college students were observed over ten weeks. Students were given a specific type of physics problem and taught a strategy that would solve the problem. During the last five weeks of the study, however, the students were given problems that could be solved using the same strategy, but the problems were from a different content area within physics. Although both the fourth grade and community college students were able to transfer the strategy from the first to the second domain, the college students made the connection more quickly.

O'Sullivan and Pressley (1984), however, attenuate the notion that strategy instruction would result in spontaneous transfer from one domain to another. Fifth grade and college students were taught problem-solving strategies and were given ample opportunity to practice. Fifth grade students required elaborated instructions in order to utilize the strategy effectively, though strategy transfer did occur for fifth grade students who received the elaborated instructions. College students also benefited more from the elaborated instruction. Furthermore, when college students were asked about what strategy they employed, only those students who received the elaborated instructions could accurately report what strategy they had used.

It appears that satisfactory transfer requires explicit instruction and deliberate practice. Fisher (1998) advises that separate instruction in domain-general and domain-specific strategies is needed. Domain-general strategy
instruction requires that the learner engage in a high degree of explicit, self-
reflective questioning about what is working and when. Even so, others insist
that explicit strategy instruction is insufficient for transfer and suggest that
achievement motivation (Garner and Alexander, 1989; Fuchs, Fuchs, Prentice,
Burch, Hamlett, Owen, and Schroeter, 2003) and ability (Carr, 1996) are critical
for successful strategy transfer.

The implementation of metacognitive knowledge studies reviewed here
included studies that employed a variety of participants, activities, and measures.
Taken together, these studies provide support for the following: 1) attempts at
metacognitive monitoring may be found in pre-school age children; 2) older
students are better at using strategies than younger students; 3) when working
on math problems, boys may be more likely to use internal strategies while girls
may be more inclined to use external strategies; 4) certain social interactions
may facilitate implementation of metacognitive knowledge; and 5) strategies
may transfer across domains under particular circumstances.

*Monitoring Between Strategies*

Strategy selection research has examined the choices made by students
after they are taught two strategies, usually with the opportunity to practice one
or both strategies prior to the testing phase. The assumption is that students will
monitor the effectiveness of each strategy during practice in conjunction with an
evaluation of the strategy and associate this information with personal
metacognitive knowledge. The research described in this section is summarized in Table 3.

In the first study to use this approach, Pressley, Levin, and Ghatala (1984) examined strategy utilization between undergraduates and fifth, sixth, and seventh graders. Students were instructed in two strategies: repetition and elaborative association. One group of students at each grade level was told that the first strategy was better than the second; another group was told that the second strategy was better than the first. Students who were simply told elaboration was better performed better. In a follow-up study, some groups of students were given the opportunity to practice both strategies and then they were allowed to choose between the two strategies. Students in these groups who chose elaboration had better recall during the testing phase. If students had to pick a strategy prior to practice, they tended to pick repetition, but did not do as well as students who chose elaboration. When fifth, sixth, and seventh grade students were given a recommendation, practice, then a choice, they followed the recommendation. The adults recognized that elaboration was a better strategy, whereas the children typically did not.

In a related study undergraduates were taught two strategies (elaboration and repetition) and then assigned to one of five groups (Pressley, Levin, and Ghatala, 1988): 1) instruction in elaboration and repetition strategies, 2) instruction and practice with the elaboration strategy, 3) instruction with the opportunity to compare both elaboration and repetition strategies during
practice, 4) practice with both strategies without instruction, or 5) instruction and practice with repetition. It was predicted that students who recognized elaboration as the better method would maintain the strategy when they were tested two weeks later. Students who were able to compare the two strategies during practice and chose elaboration did better at testing two weeks after instruction. Pressley and his colleagues argued that students who have the opportunity to try different strategies were able to identify the better strategy and used it.

A similar approach was taken with much younger students. Second grade students were assigned to one of three conditions: 1) choose the most effective strategy (strategy utility); 2) choose the more fun strategy, or 3) no directive given (Ghatala, Levin, Pressley, & Lodico, 1985). Students were then taught two strategies (repetition and elaboration) and given the opportunity to practice. Following practice, they were required to learn items from a list. At testing students were instructed to pick a strategy according to their directive. Students in the strategy utility group were the quickest to choose the more effective strategy, elaboration. Students in the strategy-utility group maintained the more effective strategy compared to students who picked the “fun” strategy when tested several weeks later. Students in the strategy-utility group indicated that the association strategy was the more effective strategy. The students in the other two groups could not identify which of the two strategies was more effective.
Table 3. Studies examining metacognitive monitoring in strategy selection.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Grade</th>
<th>Conditions</th>
<th>Results</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressley, Levin, &amp; Ghatala</td>
<td>1984</td>
<td>5, 6, 7, College</td>
<td>Strategy selection based on grade and instruction</td>
<td>Adults utilized better strategy, children obeyed experimenter directions regardless of success</td>
<td>Adults are better able to compare strategies by practice. Children did not evaluate.</td>
</tr>
<tr>
<td>Ghatala, Levin, Pressley, &amp; Lodico</td>
<td>1985</td>
<td>2</td>
<td>Strategy selection based on evaluation criteria</td>
<td>Instruction to select by effectiveness, provided better strategy selection and performance</td>
<td>Young students are capable of evaluating strategy by effectiveness with practice.</td>
</tr>
<tr>
<td>McGivern, Levin, Pressley, &amp; Ghatala</td>
<td>1990</td>
<td>2, 7, College</td>
<td>Strategy selection with a model or self-monitoring with paired associates.</td>
<td>2 grade students chose strategy at random. Other groups selected effectively either with self-monitoring or model.</td>
<td>Either self-monitoring or model are effective for seventh and college students.</td>
</tr>
<tr>
<td>Carr &amp; Jessup</td>
<td>1995</td>
<td>2</td>
<td>Strategy choice over three months in math.</td>
<td>Students moved from counters to decomposition strategies.</td>
<td>Monitoring is most important in becoming proficient in strategy implementation</td>
</tr>
</tbody>
</table>
Another tactic has been to examine different approaches to strategy instruction, with the goal of having students identify the more efficient strategy between repetition and elaboration (McGivern, Levin, Pressley, & Ghatala, 1990). Second grade, seventh grade, and college students were assigned to one of three conditions: 1) no monitoring, 2) self-monitoring, or 3) monitoring with a model. Students in the self-monitoring group were taught repetition and elaboration strategies for learning a list of paired associates. Students in the model group saw a videotape of a same sex model learning a list of paired associates using repetition, followed by the model using the elaboration strategy. All students were allowed a practice session. Second grade students were unable to accurately choose the more efficient strategy in any condition. Seventh grade and college students did equally well in the self-monitor and the model groups. The college students in the model group performed slightly better than the second and seventh grades students, though not significantly.

One study limited strategy instruction to the domain of math. Second grade students received instruction for three strategies to be used in solving addition and subtraction problems: 1) retrieval, 2) counters, and 3) decomposition (breaking addends into tens and ones) (Carr & Jessup, 1995). Strategy selection, metacognitive knowledge, and metacognitive ability were measured in April and then again in June. Monitoring occurred when the task was challenging and the process was not yet automatized (i.e., used retrieval or a less demanding strategy). Students tended to migrate from the counters
strategy in April to the decomposition strategy by June. Students who used retrieval methods for solving the math problems in April tended to rely on retrieval in June. Metacognitive monitoring was highest in June in association with the decomposition strategy. The researchers suggested that metacognitive monitoring was more important when a student was less proficient with the use of a strategy. It may be that as proficiency improves so does monitoring.

To summarize, older students were more likely to compare strategy effectiveness and pick the appropriate strategy, while younger students tended to do what they were instructed to do. Even so, older students require practice with the strategies in order to select the more efficient strategy. Younger students could not choose the more effective strategy through observation or self-monitoring. When instructed in complex strategies, young students required several months to adapt to the new strategy.

*Errors in Monitoring*

A different approach to the study of metacognitive monitoring is to examine the failure to monitor progress at a task. In a study of this type, students are typically given a passage to read with the instruction to identify inconsistencies within the text. The research reviewed in this section is summarized in Table 4.

Baker (1979) reported that students employ little monitoring during reading when they do not identify inconsistencies in a text. College students were given texts to read with the instructions to identify inconsistencies inserted
into the text and describe how these inconsistencies affect comprehension.

There were three types of inconsistencies: 1) inconsistent information (a statement that was inconsistent with the main idea), 2) unclear reference (an ambiguous statement in relation to the main idea), and illogical connection (a statement that conflicted with the main idea). The students did not identify 62% of the inconsistencies. Inconsistencies that were identified tended to be categorized as "inconsistent information" or "unclear references." It was ascertained from interviewing the students after testing that students employed "fix-up" strategies to maintain comprehension. For example, students assumed insufficient information was in the text to resolve the inconsistencies or they used personal background knowledge to resolve the inconsistencies.

Markman (1979) also examined errors in reading comprehension with third, fifth, and sixth grade students. He proposed that in order for students to identify inconsistencies explicit standards about what constitutes an inconsistency must be presented. Even with this effort, third grade students could not identify most inconsistencies. Fifth and sixth grade students identified some of the inconsistencies when told they were in the text. The authors concluded that success at identification of the inconsistencies required a heavy cognitive toll on students. Students needed to encode the information, draw out the relevant inferences, and maintain the inferences in working memory while reading the material. The heavy demands of this task required that students be...
Table 4. Studies examining errors in metacognitive monitoring.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Grade</th>
<th>Conditions</th>
<th>Results</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baker</td>
<td>1979</td>
<td>College</td>
<td>Detect inconsistencies in text</td>
<td>Many inconsistencies not identified.</td>
<td>College students may apply fix up strategies as compensation.</td>
</tr>
<tr>
<td>Markman</td>
<td>1979</td>
<td>3, 5, 6</td>
<td>Detect inconsistencies in text after explicit instruction</td>
<td>Students could not identify inconsistencies with instruction</td>
<td>Heavy processing toll prevents monitoring.</td>
</tr>
<tr>
<td>Baker &amp; Anderson</td>
<td>1982</td>
<td>College</td>
<td>Detect inconsistencies with directions and rereading text</td>
<td>Higher error detection</td>
<td>Informing college students of types of inconsistencies made small improvement in performance</td>
</tr>
<tr>
<td>Pressley, Ghatala, Pirie, &amp; Woloshyn</td>
<td>1990</td>
<td>College</td>
<td>Identify main idea by rereading or be highly certain.</td>
<td>High confidence group did not perform better than control group.</td>
<td>Directions lead to overconfidence in identifying main idea.</td>
</tr>
</tbody>
</table>
instructed to look for inconsistencies. The authors proposed that elementary grade students do not spontaneously engage in monitoring. Baker and Anderson (1982) gave college students three texts to read that either: 1) contained information that was inconsistent with the main point, 2) contained details within the passage that were inconsistent, or 3) had no inconsistencies with the gist of the passage. Half of the students were told some texts would contain inconsistencies. Subjects were encouraged to reread sections of the text. Measures of reading comprehension consisted of reading time and of the number of inconsistencies identified. Telling students there would be inconsistencies had a small effect on reading performance. Sixty-six percent of the main point inconsistencies were identified and comparable performance was observed with respect to identifying detail inconsistencies. Forty-nine percent of the students, however, failed to identify one or both types of inconsistencies. Students spent more time on the inconsistent aspects of the text than on sections of the text that were consistent with main point.

In summarizing her work, Baker (1989) observed that adult readers tend to have greater metacognitive awareness and more expertise in monitoring compared to elementary school age children. Adult readers, however, do not monitor their own comprehension well. Adults' perception of competency in comprehension was much higher than actual competency when tested. Baker suggested that adult readers still have plenty of room for improvement in monitoring reading comprehension. While spontaneous strategy use has been
identified around fifth or sixth grade (Kuhn, 2000b), most adults have not mastered the complexities of metacognitive monitoring (Baker, 1979; Pressley, Ghatala, Pirie, & Woloshyn, 1990).

**Monitoring Declarative Knowledge**

The final approach in examining metacognitive monitoring is to have participants make confidence judgments about how accurately they can produce an answer to content to which they have had previous exposure. In studies of this nature, students may make predictions about an item they are about to attempt or judge how accurately they have answered an item. Judgments may be on an item-by-item basis or cover an entire set of items. This focus on the monitoring of declarative knowledge has lead to the identification of domain-general and domain-specific metacognitive monitoring. The research described in this section is summarized in Table 5.

Bisanz and colleagues compared differences in elementary and college students' monitoring (Bisanz, Vesonder, & Voss, 1978). Students in first, third, and fifth grades were compared to each other and with college students in monitoring on recall in a paired associate task. After each pair was recalled, students were asked to judge the accuracy of their recall. First grade students were not very accurate and reported a significant percentage of false positives. Fifth grade students' discriminated fairly well between correct versus incorrect responses. Fifth grade students, however, were not as accurate as the college students.
<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Grade</th>
<th>Conditions</th>
<th>Results</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisanz, Vesonder, &amp; Voss 1978</td>
<td>1, 3, 5, college</td>
<td>Confidence in recalling word pairs correctly.</td>
<td>Older students better at discriminating between what is known and what is not known.</td>
<td>Accuracy in confidence judgments improve with age</td>
</tr>
<tr>
<td>Schneider, Korkel, &amp; Weinert 1989</td>
<td>3, 5, 7</td>
<td>Feeling-of-knowing judgments about soccer knowledge</td>
<td>Relied on knowledge when available, otherwise relied on domain-general strategies to answer</td>
<td>Relied on domain-general strategy when lacking domain-specific knowledge</td>
</tr>
<tr>
<td>Schraw, Dunkle, Bendixen, &amp; Roedel 1995</td>
<td>College</td>
<td>Confidence judgments across different domains</td>
<td>Domain-specific and domain-general monitoring identified in correlations and principal component analysis.</td>
<td>Domain-specific and domain-general processing may both occur and contribute to monitoring.</td>
</tr>
<tr>
<td>Schraw 1997</td>
<td>College</td>
<td>Confidence ratings in four domains</td>
<td>Identified domain-general factor</td>
<td>Domain domain-general processing present.</td>
</tr>
<tr>
<td>Schraw &amp; Nietfeld 1998</td>
<td>College</td>
<td>Monitoring of fluid and crystallized ability tasks.</td>
<td>Crystallized ability and fluid ability identified in correlations and PCA</td>
<td>Monitoring may be relevant to type of task.</td>
</tr>
<tr>
<td>Keleman, Weaver, &amp; Epstein 2000</td>
<td>College</td>
<td></td>
<td>Found no correlations between types of tasks. Refutes transfer hypothesis of domain-general monitoring.</td>
<td></td>
</tr>
<tr>
<td>Kletman &amp; Stankov 2001</td>
<td>College</td>
<td>Confidence ratings on domain-general, domain-specific, and perception measures.</td>
<td>High correlations between and within tests. Factor analysis identified domain-specific and domain-general monitoring.</td>
<td>Task and individual differences influence monitoring</td>
</tr>
<tr>
<td>Rozencwajg 2003</td>
<td>College</td>
<td>Confidence ratings on fluid and crystallized tasks.</td>
<td>Metacognitive knowledge related to fluid ability, metacognitive monitoring</td>
<td>Monitoring measures may be related to domain-general processing.</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Grade</td>
<td>Conditions</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Weaver &amp; Keleman</td>
<td>2003</td>
<td>College</td>
<td>Evaluate different types of metacognitive judgments</td>
<td>Judgments were correlated across two trials, but not across different types of metacognitive judgments.</td>
</tr>
</tbody>
</table>
Schneider, Korkel, and Weinert (1989) examined the strategies that third, fifth, and seventh grade students used when answering domain-specific questions. Schneider and colleagues tested all students on their knowledge about soccer. Students also provided feeling-of-knowing judgments. That is, when students answered a question incorrectly they were asked to judge the likelihood of selecting the correct answer from a list of options. Students were divided into two groups: those with knowledge about soccer and those who without knowledge about soccer. Seventh grade students had more accurate feeling-of-knowing judgments than third grade students. Students with domain knowledge relied on that personal knowledge to answer test items; whereas students without specific knowledge relied on domain-general strategies to answer test items.

In the first explicit examination of domain-general metacognitive monitoring, Schraw, Dunkle, Bendixen, and Roedel (1995) assessed monitoring with eight domain multiple-choice tests: 1) geography, 2) American history, 3) caloric value of foods, 4) running speeds of animals, 5) mathematical word problems, 6) spatial judgments, 7) general knowledge questions, and 8) the Nelson-Denny reading comprehension test. Participants gave confidence ratings for each item on each test. Four types of scores were included in the analysis: 1) performance—the raw performance scores; 2) confidence—the raw ranking of how confident they felt about each answer; 3) discrimination—a measurement of how well the confidence rating matched actual performance; and 4) bias—a
measurement of overconfidence when compared to actual performance. Low positive correlations in performance scores and discrimination scores between domains were considered indicative of domain-specificity. Schraw and colleagues suggested that the high positive correlation found between confidence and bias scores across all domains supported domain-general metacognitive monitoring; metacognitive monitoring that is independent of domain-specific knowledge. The strong, positive correlations across tasks in monitoring judgments suggested domain-general metacognitive monitoring was being employed, whereas the low correlations suggested employment of domain-specific metacognitive monitoring. Each type of score was analyzed separately using principal component analysis. Performance and discrimination scores produced several components, which were interpreted as domain-specific metacognitive monitoring factors, while confidence and bias scores produced a single component, which was interpreted as a domain-general metacognitive monitoring factor. In a follow-up study, five tests were given to participants (presidential history, geography, vocabulary, music history, and a sports test) (Schraw, Dunkle, Bendixen, and Roedel, 1995). Items were matched on difficulty, format, and type of inferences necessary for providing a response. The four types of scores (performance, confidence, accuracy, and bias) were again examined and strong, positive correlations were observed between all domains. One principal component was extracted that reflected all scores. While the principal component analysis suggested the presence of domain-general metacognitive monitoring when the domains were
matched on difficulty and format, the low correlations between domains suggested the presence of domain-specific metacognitive monitoring.

During further examination of domain-general metacognitive monitoring (Schraw, 1997) participants were tested across several domains including: 1) lexical word identification, 2) reading comprehension, 3) syllogistic reasoning, and 4) mathematical reasoning. Significant, positive correlations on confidence ratings between domains were observed which was interpreted as support for the presence of domain-general metacognitive monitoring. Subjects also completed a survey about general monitoring strategies. Confidence scores in each domain correlated highly with the monitoring survey. Performance scores, on the other hand, did not. The results were interpreted to suggest that domain specific metacognitive monitoring was reflected in the low, positive correlations between performance scores; while domain-general metacognitive monitoring was marked by significant positive correlations between confidence judgments.

Schraw and Nietfeld (1998) redefined domain-general and domain-specific metacognitive monitoring in terms of fluid ability (i.e., monitoring not related to specific content knowledge) and crystallized ability (i.e., monitoring related to specific content knowledge). Participants (university undergraduates) took the following three tests reflecting fluid ability: 1) the Raven’s Progressive Matrices, 2) the Schaie-Thurstone Letter Series, and 3) and a probabilistic reasoning task. Five additional tests were used to represent crystallized ability: 1) reading comprehension, 2) vocabulary knowledge, 3) American history, 4)
geography, 5) and music history. Confidence ratings were obtained for each item on every test. Bias, an overconfidence measure, and monitoring accuracy scores were collected. Significant positive correlations between tests that measured fluid ability were observed. Significant positive correlations also occurred between tests that measured crystallized ability. A principal component analysis using bias scores and accuracy scores resulted in two components: one for crystallized ability, or domain-specific metacognitive monitoring, and one for fluid ability, or domain-general metacognitive monitoring. A principal component analysis using the performance scores resulted in three components: 1) verbal crystallized ability, 2) general information crystallized ability, and 3) fluid ability.

Additional support for the existence of domain-general metacognitive monitoring comes from a study by Kleitman and Stankov (2001). In this study, participants were tested on geography questions, the Raven's Progressive Matrices Test, and a line length test. The Raven's Matrices was used as an indicator of fluid ability, the geography test was used to measure crystallized knowledge, and the line length test was used as a perception measure. Average confidence ratings were highest for the geography test (85%), followed by the Raven's Progressive Matrices (60%), and finally the line length test (44%). In terms of bias scores, or degree of overconfidence, with respect to the line length test participants were underconfident, whereas for the other two tests participants were slightly overconfident. Strong, positive correlations on bias scores between all tasks were observed. A principal component analysis resulted
in three monitoring components: one for each domain-specific task (i.e., one for geography, one for the line length test, one for Raven's), and a domain-general confidence component.

Kleitman and Stankov divided confidence judgments into two types: during the task (monitoring) and after the task (evaluation). Kleitman and Stankov suggested that metacognitive monitoring judgments were influenced by both the task and individual differences.

Rozencwajg (2003) examined problem solving in terms of crystallized and fluid ability. College students completed physics problems, math problems, and a sentence completion task. In addition, students completed a task that required identification of a missing cell from a matrix, a task similar to the Raven's Progressive Matrices, as a measure of fluid ability. Performance on physics problems was positively correlated with both fluid ability (the matrix test) and crystallized ability (sentence completion and math problems). Metacognitive knowledge was strongly, positively correlated with crystallized ability; whereas metacognitive monitoring was strongly correlated with fluid ability. Metacognitive knowledge and metacognitive monitoring shared a moderate, positive correlation. There was a strong, positive correlation between metacognitive monitoring and fluid ability, but the correlation between metacognitive monitoring and crystallized ability was rather weak. Metacognitive knowledge and crystallized ability shared a strong, positive correlation but fluid ability and metacognitive knowledge were not correlated. The partial correlation between
metacognitive monitoring and crystallized ability (holding fluid ability constant) was approximately zero. When fluid ability was correlated with monitoring and crystallized ability was controlled, however, the partial correlation remained strong and positive. The results were reversed for metacognitive knowledge. When crystallized ability was controlled, the partial correlation between metacognitive knowledge and fluid ability was near zero. When fluid ability was controlled, the partial correlation between metacognitive knowledge and crystallized ability remained moderate.

The implications of these studies are that two separate processes, domain-general and domain-specific metacognitive monitoring, were established, as was their connection to metacognitive activity in general. Crystallized ability appears to reflect domain-specific knowledge as well as specific metacognitive strategies. Fluid ability reflects domain-general metacognitive monitoring in circumstances when specific knowledge is not available or not relevant. In addition, metacognitive monitoring may be more applicable to fluid ability tasks.

Kelemen, Weaver, and Epstein (2000) asserted that there is no such thing as domain-general metacognitive monitoring. In their study, four types of metacognitive monitoring judgments were examined over two trials: 1) ease-of-learning judgments (i.e., how difficult would each item be to learn in a paired-associates task?), 2) judgments-of-learning (i.e., how well have you learned a particular item in a paired associates task?), 3) feeling-of-knowing judgments (i.e., if you cannot recall the item, how confident are you that you could identify
the correct answer in a recognition task?) and 4), text comprehension monitoring. Ease-of-learning judgments and judgments-of learning were made when participants learned lists of word pairs, whereas feeling-of-knowing judgments involved having a participant judge how likely they thought it would be that they would answer a general knowledge question correctly (such as identifying the tallest mountain in South America). The text comprehension judgments consisted of having participants rate their understanding of a passage they had read. Correlations between the four different types of metacognitive judgments were not significant; however, individual types of metacognitive judgments were strongly, positively correlated between the two trials.

In an attempt to shed some light on the different outcomes between Schraw's (1997) work and Kelemen's work (Kelemen, Weaver, & Epstein 2000), Campbell and Guadagnoli (2004) combined methodologies from both lines of study. In this study, participants produced three confidence judgments (ease-of-learning judgments, judgments-of-learning, and feeling-of-knowing judgments) in a paired associates task and a general knowledge test over two trials one week apart. Significant, positive correlations were observed between judgments-of-learning and feeling-of-knowing judgments. Based on this study, Campbell and Guadagnoli concluded that the different outcomes between Schraw's (1997) and Kelemen's work (Kelemen, Weaver, & Epstein, 2000) were artifacts of methodological approaches. Kelemen's weak correlations were related
to separate tasks and not indicative of an absence of domain-general metacognitive monitoring.

Based on the studies described in this section, four main conclusions can be drawn about monitoring declarative knowledge: 1) there is wide agreement that individuals are accurate metacognitive monitors within specific content areas, 2) domain-general metacognitive monitoring may be engaged for novel tasks or for problem-solving tasks, 3) older students are more accurate metacognitive monitors than younger students, and 4) students may rely on domain-general monitoring when domain-specific information is not available.

Two hypotheses have been put forth related to domain-general metacognitive monitoring. The first is that domain-general metacognitive monitoring is employed in novel tasks when one cannot rely on specific content knowledge (Schneider, Korkel, & Weinert, 1989). The second is that domain-general metacognitive monitoring is the result of transfer of metacognitive monitoring skills in a domain specific sense to other domains (Schraw & Nietfeld, 1998). These two hypotheses will be examined in greater depth later in this review.

Working Memory

Many have classified metacognitive monitoring as a central executive function within working memory (other central executive functions include attention, planning, and evaluation) (Nelson & Narens, 1990; Nelson, 1996;
Rozencwaig, 2003). Like metacognitive monitoring, working memory has been separated into general and specific processes. The central executive represents a general process, while the phonological loop and visuospatial sketchpad have been hypothesized to be separate and highly specialized information processing subsystems (Baddeley & Hitch, 1974; Baddeley, 2002; Swanson, 1996; 1999).

Bayliss and colleagues (Bayliss, Jarrold, Gunn, & Baddeley, 2003) suggested that working memory consists of a domain-general processing component appropriate for information that has specific storage sites. A study put forth to support this view required that third and fourth grade students and college undergraduates complete a series of working memory storage (digit span, Corsi span) and processing tasks (verbal association task and a visual search task) in addition to reading and mathematics achievement tests. The data analysis supported the hypothesis that the two specialized storage systems, the visuospatial sketchpad and the phonological loop, made separate contributions to performance, as did processing from the central executive. Bayliss and colleagues suggested that domain-general metacognitive monitoring might function within working memory domain-general processes (Pressley & Ghatala, 1990; Bayliss, Jarrold, Gunn, & Baddeley, 2003). Combining this information with the current central executive model could lead to the impression that domain-specific metacognitive monitoring focuses on domain content knowledge.

Similar separation of domain-specific and domain-general processes has been provided by Swanson' examination of working memory functions with a
large range of children (1999; Swanson, 1992; Swanson, 2004). Swanson compared verbal and visuospatial working memory tasks with reading and mathematics tasks and found that age-related changes appeared to be the result of domain-general processing and not domain-specific processing. In other words, working memory processing differences between ages appears to be related to changes in working memory capacity and not to more efficient processing.

Carr and colleagues (Carr, Alexander, and Folds-Bennett, 1994) have suggested that metacognitive monitoring occurs in young children for tasks that do not overtax working memory. In one study, second grade students were surveyed about mathematical strategy knowledge in September and then again in January. Use of mathematics strategies (e.g., external counters or solving the problem internally) and attributions of effort were assessed through interviews with the students. Most students possessed mathematics metacognitive knowledge at the beginning of second grade. Students with higher metacognitive knowledge were more likely to use internal strategies, such as memorization, rather than external strategies, such as the use of counters or fingers. Students with higher metacognitive knowledge attributed their success in math to personal effort rather than ability. Most students showed an overall increase of internal strategies by January. Metacognitive knowledge and effort attributions also increased. Given the Swanson and Carr studies, it is clear that any attempt to
study the development of metacognitive monitoring across several years must account for changes in working memory.

In sum, metacognitive monitoring is likely a working memory process. Any hypothesis about domain-general and domain-specific monitoring must coincide with current knowledge related to working memory. Research supports the model of the central executive as a single general process, while the phonological loop and visuospatial sketchpad appear to be modular processing systems (Baddeley & Hitch, 1974; Baddeley, 2002; Swanson, 1996; 1999). Metacognitive monitoring should not occur if working memory is overtaxed. Domain-specific metacognitive monitoring may be influenced by specific storage, whereas domain-general metacognitive monitoring may relate more specifically to the general executive working memory process.

**Summary and Conclusions**

From this literature review the conclusion may be drawn that metacognitive monitoring has been associated with superior performance on a variety of tasks. Better metacognitive monitoring and better performance have been associated with increases in metacognitive knowledge and strategy practice. Metacognitive monitoring is more likely to occur with moderately difficult tasks or when an individual cannot rely on specific knowledge. If the task is overly demanding, neither domain-general nor domain-specific metacognitive
monitoring may occur. If working memory is overtaxed, effective metacognitive monitoring is unlikely.

Domain-specific metacognitive monitoring refers to specific regulation of cognitive processes related to knowledge within specific domains. Without adequate metacognitive monitoring, reading comprehension and problem solving are hindered. Specific instruction in metacognitive knowledge and strategy skills can improve reading comprehension and problem solving skill.

Domain-general metacognitive monitoring has been defined as monitoring that does not rely on specific content knowledge (English, 1992; Schneider, 1998). It has also been hypothesized that domain-general metacognitive monitoring may span domains of knowledge (Gustafsson & Undheim, 1996; Schraw, Dunkle, Bendixen, Roedel, 1995; Veenman & Verheij, 2003), though studies of strategy transfer are inconclusive. Strategy transfer can be taught (Kuhn 1995; O'Sullivan & Pressley, 1984), but transfer across domains typically does not occur spontaneously (Ghatala, Levin, Pressley, & Lodico, 1985; Ghatala, 1986).

Several themes can be derived from the literature. The first theme is that older students are better metacognitive monitors than younger students. Metacognitive monitoring has been identified in children as young as four-years old (Cultice, Somerville, & Wellman, 1982; English, 1992; Fletcher-Flinn & Snelson, 1997). Pre-school children have demonstrated basic metacognitive understanding and simple metacognitive monitoring skills. Strategy use
monitoring becomes more efficient with maturation. The development of metacognitive monitoring is considered linear, but slow. The ability to reflect on metacognitive knowledge and monitoring also improves with maturation. More complex monitoring skills require explicit instruction and substantial practice (Garner and Alexander, 1989; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, and Schroeter, 2003).

A more specific look at the developmental trends for domain-specific and domain-general metacognitive monitoring, leads to the view that practice is the primary reason for improvement. This improvement includes more efficient metacognitive monitoring and the implementation of more complex strategies. Domain-specific metacognitive monitoring is engaged for work within specific content areas (Armbruster, Echols, & Brown, 1982; Cross & Paris, 1988; Schneider, Korkel, Weinert, 1989; Carr, Alexander, & Folds-Bennett, 1994; Carr & Jessup, 1995). Students may acquire substantial practice with a well-developed curriculum. Students proficient in specific content monitoring skills should perform better in domain-specific monitoring than domain-general metacognitive monitoring.

The second theme that may be identified is that metacognitive monitoring benefits from instruction. Explicit instruction in domain-specific metacognitive monitoring has been positive (Armbruster, Echols, & Brown, 1982; Carr, Alexander, & Folds-Bennett, 1994; Carr & Jessup, 1995; Cross & Paris, 1988) In order to achieve measurable effects, however, explicit metacognitive monitoring
instruction was required in addition to several weeks of practice (Brown & Smiley, 1978; Palinscar & Brown, 1984). Domain-specific metacognitive monitoring may not be an ability that gradually develops over time or on its own. This is why some have called for more metacognitive instruction (Hall, Myers and Bowman, 1999). Domain-general metacognitive monitoring is not taught explicitly in schools, even though domain-general strategies—like elaboration—have been successfully taught and used (Ghatala, Levin, Pressley, & Lodico, 1985; Pressley, Ghatala, & Levin, 1988; McGivern, Levin, Pressley, & Ghatala, 1990).

The third theme is that task, such as difficulty, or personal, such as age, attributes may interfere with the ability to monitor during a task (Campbell & Corkill, 2004; Carr, Alexander, and Folds-Bennett, 1994). Difficult tasks tax available cognitive resources, which may prevent the use of newly acquired strategies. In addition, monitoring of strategy effectiveness can only occur with sufficient practice (Palinscar & Brown, 1984; Garner and Alexander, 1989). It may be that practice leads to automatization when domain-specific knowledge is available or if the individual has reached expertise with a set of skills. Finally, successful strategy transfer also requires explicit instruction and opportunities for practice (Fisher, 1998; Ghatala, 1986; Weaver & Keleman, 2003).

Two hypotheses relating to the development of domain-general metacognitive monitoring have been proposed. The transfer-appropriate hypothesis (Schraw & Nietfeld, 1998; Weaver & Keleman, 2003) requires that
domain-general metacognitive monitoring develop after domain-specific metacognitive monitoring. This sequence is hypothesized to be the result of transfer of metacognitive knowledge from one domain to another. Support for this hypothesis comes from studies involving factor analysis (Schraw & Nietfeld, 1998) and some transfer studies (Kuhn, 1995; Fisher, 1998). Others have refuted this hypothesis (Ghatala, 1986; Keleman & Weaver, & Epstein, 1997; Weaver & Keleman, 2003).

The other hypothesis is a default hypothesis in which domain-general metacognitive monitoring is the default process until domain expertise and domain-specific metacognitive monitoring develop. Research supporting this hypothesis comes from studies that show that children use on domain-general strategies when domain-specific knowledge is absent (Schneider, Korkel, and Weinert, 1989; English, 1992; Short, Schatschneider, & Friebert, 1993). Neither developmental hypothesis has been explicitly tested.

The Present Study

The purpose of the present study was to identify the developmental trend of domain-specific and domain-general metacognitive monitoring through childhood. An additional goal was to identify the relationship, if any, between domain-general and domain-specific metacognitive monitoring. Although researchers have measured metacognitive monitoring in childhood for both domain-specific and domain-general tasks (e.g., English, 1992; Schneider,
Korkel, & Weinert, 1989; Ghatala, 1985), no studies have considered the development of domain-specific and domain-general metacognitive monitoring simultaneously. A concurrent examination of the development of domain-general and domain-specific metacognitive monitoring is needed.

Domain-specific metacognitive monitoring is thought to result from explicit strategy instruction and practice (e.g., Brown, Palinscar, & Armbruster, 1984; Paris & Jacobs, 1984; Lucia, 1994; Mevarech & Kramarski, 2003). Domain-general metacognitive monitoring is thought to be employed when an individual engages in novel tasks (English, 1992; Schneider, Korkel, & Weinert, 1989; Weaver & Keleman, 2003). Individuals should be better monitors (more accurate) when employing domain-specific metacognitive monitoring because domain-general metacognitive monitoring should not benefit from content-specific strategy instruction and practice. Therefore, it was hypothesized that greater accuracy would be seen under circumstances that called for domain-specific, rather than domain-general metacognitive monitoring.

Currently there is no theoretical explanation for the development of domain-specific and domain-general metacognitive monitoring. The findings of this study may allow for better understanding of metacognitive monitoring and may assist in the development of metacognitive strategy instruction. The findings of this study may provide information relevant to strategy instruction within school curricula. If, for example, domain-general strategies are built out of
domain-specific strategies, then strategy instruction should focus on domain-specific strategies.

**Hypotheses**

Two hypotheses have been identified as routes for the development of domain-specific and domain-general metacognitive monitoring. Schraw and Nietfeld (1998) proposed an experiential hypothesis. This hypothesis is that metacognitive monitoring develops through practice and instruction. The second possibility is dubbed the modularization hypothesis. Modularization theorists would propose that brain processes move from general to specific as a result of experience (Karmiloff-Smith, 1992; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Karmiloff-Smith, Plunkett, Johnson, Elman, & Bates, 1998). Therefore, metacognitive monitoring should follow suit.

According to the experiential hypothesis, students become more proficient in content knowledge and strategy knowledge through practice and instruction (Schraw & Nietfeld, 1998). Metacognitive monitoring begins as a domain-specific process then at some later point generalizes into a domain-general process. These theorists would propose that domain-specific metacognitive monitoring would appear in younger students with domain-general metacognitive monitoring not appearing until later in childhood (see Figure 1). Currently, there is no research indicating when or where this generalization would occur. The present study, therefore, would be an initial attempt to pinpoint the developmental
timeframe. The experiential hypothesis would be supported by weak relationships between domain-general or domain-specific tasks in earlier grades, but stronger and positive correlations among domain-general and domain-specific tasks in later grades. The presence or absence of significant correlations has been used in the past as evidence for the presence of domain-general metacognitive monitoring.

Monitoring studies, in general, do not support the later development of domain-general processing. Several studies (e.g., English, 1992; Schneider, Korkel, & Weinert, 1989) have documented domain-general metacognitive monitoring in children as young as four years old. Schneider, Korkel, and Weinert (1989), for example, suggested that domain-general metacognitive monitoring is the default process for novel situations. Ghatala (1986) and Weaver and Keleman (2003) did not observe spontaneous strategy transfer from one domain to a second domain. The experiential hypothesis would require that domain-general metacognitive monitoring develop as the result of the transfer of metacognitive knowledge and strategy use across content domains. Researchers who have studied strategy transfer detected it mainly in instances when explicit instruction was provided and/or deliberate practice was required (Kuhn, 1995; Fisher, 1998). This leads to the impression that domain-general metacognitive monitoring would require explicit instruction and practice. However, as previously noted, domain-general metacognitive monitoring does occur without explicit instruction (Schneider, Korkel, & Weinert, 1989), practice, and perhaps without
transfer of metacognitive and strategy knowledge. Finally, the experiential hypothesis requires the development of two processes within working memory: one for domain-specific metacognitive monitoring and one for a domain-general process. A two-process metacognitive monitoring model conflicts with current conceptions of working memory (Nelson, 1996; Baddeley, 2002; Baddeley, 2002; Swanson, 1996).

Figure 1. Expected developmental progression according to the experiential hypothesis.

The second possibility, the modularization hypothesis, would require that metacognitive monitoring first be observed as a domain-general process that becomes domain-specific as a result of experience (see Figure 2). This second hypothesis stems from neural network models (Karmiloff-Smith, 1992; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Karmiloff-Smith, Plunkett, Johnson, Elman, & Bates, 1998). Based on the language acquisition
model of Karmiloff-Smith (1992; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996), brain development follows general predispositions, but becomes specialized, or modularized, with exposure to specific experiences. Proponents of the modularization hypothesis would suggest that there is a continuum of metacognitive monitoring on which individuals move from domain-general to domain-specific with experience and practice. Consider, for example, reading instruction. Reading instruction typically occurs very early in elementary school. Specific knowledge and strategy skills are taught and practiced until competence is achieved. Metacognitive monitoring in reading is better in later than earlier grades due to practice and experience. Delclos and Harrington (1991), for example, investigated reading and domain-specific metacognitive monitoring in students in third, fifth, seventh, eleventh and twelfth grade, and college. Students at each grade level performed better than the grade level below up to and including the high school age students. Other researchers (e.g., Pressley, Levin & Ghatala, 1984; Schneider, 1998) have suggested that by late high school, metacognitive monitoring has evolved into an automatic process. The current study extends the research on the development of domain-specific metacognitive monitoring by contrasting that developmental path with the development of domain-general metacognitive monitoring.

According to the modularization hypothesis, domain-general metacognitive monitoring should be present in early childhood as a default process. Evidence for this hypothesis would be strong, positive correlations between domain-
specific and domain-general tasks, specifically in the earlier grades. Studies indicate the presence of domain-specific and domain-general metacognitive monitoring throughout the various levels of education (e.g., Brown & Smiley, 1978; English, 1992; Short, Schatschneider, & Friebert, 1993; O'Sullivan & Joy, 1994). Additional support for the modularization hypotheses would be evidence that students demonstrate greater accuracy when engaged in domain general, as opposed to domain specific, metacognitive monitoring. The modularization hypothesis fits well with current conceptualizations of working memory (e.g., Schneider et al, 1989).

Predictions
First, it is predicted that the youngest students will employ domain-general metacognitive monitoring as evidenced by significant positive correlations across the four tasks (two domain-general and two domain-specific) because developmental studies suggest that domain-general metacognitive monitoring is the default strategy. For the older students, there should not be significant correlations between the domain-specific tasks nor between the domain-specific and domain-general tasks.

Second, older students should be more proficient when engaged in domain-specific metacognitive monitoring for the tasks than the younger students. Research suggests domain-specific metacognitive monitoring develops and improves from explicit instruction, content knowledge, and practice within a specific domain. Therefore one would expect that domain specific metacognitive monitoring would improve with age. There should be no differences in ability to engage in domain-general metacognitive monitoring for the tasks between students in different grades. If it is the case that domain-general tasks do not benefit from practice, then performance domain-general metacognitive monitoring for tasks will improve only as a function of maturation. To account for practice effects, maturation will be controlled, using working memory tasks as covariates. By controlling for working memory, domain-general metacognitive monitoring should remain stable across student ages, while metacognitive monitoring for domain-specific tasks should improve (see Figure 2). These predictions, if validated, would support the modularization hypothesis.
CHAPTER 3

METHODOLOGY

Participants

Students were recruited from first, fourth, and seventh grades in this cross-sectional study. Thirty students were recruited from the first and fourth grades with 31 students from seventh grade. The total number of participants was 91. Subjects with special educational requirements or potential language barriers were excused from the research.

Materials

Two domains were chosen for measuring domain-specific metacognitive monitoring: arithmetic and vocabulary. Forty items from the arithmetic section of Wide Range Achievement Test, Third Edition (WRAT3) (1993) were used as the arithmetic measure. First grade students began this task with oral arithmetic (counting) then moved onto the written section. The fourth and seventh grade students were exposed only the written arithmetic section. The arithmetic section covered content from addition (e.g., $2 + 7 = ?$) to algebra (e.g., find $f(-2)$ where $f(x)=3x^2+x-7$). The test was designed to take no more than 15 minutes. The test ended after the participant made ten consecutive errors. Confidence
judgments for the last ten incorrect responses were included in the calculation of the metacognitive monitoring accuracy scores.

The reading section of Wide Range Achievement Test, Third Edition (WRAT3) (1993) was used for the vocabulary test. The WRAT3 measures reading achievement from ages five to seventy-four. The first grade students began with the letter identification task and then moved on to a word pronunciation task. The fourth and seventh grade students began with the word pronunciation task. The word list ranged from "in" to "terpsichorean". This test was designed to take no more than 15 minutes. Like the mathematics test, this test ended after the participant made ten consecutive errors. Confidence judgments for the last ten incorrect responses were included in the calculation of the metacognitive monitoring accuracy scores.

Domain-general, or fluid, tasks have been associated with reasoning tasks that do not rely on specific content knowledge, rather they consist of assembly and control problems; in other words, required evaluation of relationships and testing personal hypotheses (Carpenter, Just & Shell, 1990; Schraw & Nietfeld, 1998; Marshalek, Lohman, & Snow, 1983). Tasks that required strategy shifting have also been classified as domain-general (Marshalek, Lohman, & Snow, 1983). Two tasks were identified for use in this study: the Raven’s Standard Progressive Matrices Test and the Wisconsin Card Sort Task (WCST). The Raven’s Progressive Matrices Test has been associated with fluid ability (Carpenter, Just & Shell, 1990; Schraw, Dunkle, Bendixen, Roedel, 1995; Schraw
& Nietfeld, 1998), as has the WCST (Golden, Kushner, Lee, McMorrow, 1998; Laws, 1999; O'Donnell, MacGregor, Dabraowski, Oestreicher, & Romero, 1994). Both provide norm tables across a large span of ages allowing for comparisons between students at various grades levels.

In the Raven's Standard Progressive Matrices Test participants were shown geometric patterns in a 3 x 3 matrix. The bottom right cell was empty. Participants were expected to choose the most appropriate match from four choices. This test is considered appropriate for participants between the ages of six and eighty. The test consisted of 64 items. The experimenter hand scored each test for number correct. Scores were coded as percentile ranks.

The WCST is appropriate for use with individuals between age six and eighty-nine. The WCST consists of four stimulus cards and 128 response cards. Each card contains three characteristics: 1) color (blue, yellow, green, and red), 2) form (circle, star, cross, and triangle), and 3) number (1, 2, 3 or 4 items). For example, a card might have two, blue stars or one, red triangle. To administer the test, the researcher lays out four stimulus cards. These cards are one red triangle, two green stars, three yellow crosses, and four blue dots. The participant is given a deck of 128 cards. The participant looks at the top card from the response deck and matches that card to one of the stimulus cards based on whatever they wish; they do not know the sorting criteria. The experimenter then tells the participant if they have matched the card correctly. The participant then sorts the next card in the stack and is told whether that sort...
is correct. The participant sorts cards until they deduce the correct sorting criteria and then they sort 10 cards based on that criteria. After ten cards are sorted correctly, the sorting criteria changes without notice and the participant must adjust and deduce the new sorting criteria. The participant is required to match ten number cards first. Once this is achieved, the participant has to sort by color and then the criterion is changed to form. The process is repeated and the criterion changes from number, to color, to form after every ten consecutive correct matches. This test typically lasts 20 minutes. The experimenter hand scored the test for number correct. Scores were coded as percentile ranks.

In an effort to account for maturation, two working memory tasks were included for use as potential covariates. The first was a modified form of the sentence span task created by Daneman and Carpenter (1980). This task has been modified for children between the ages of five and nineteen by Swanson (1996). In this task, the participant listens to a set of several sentences. After the presentation of the set of sentences, the participant is required to recall the last word of each sentence in the order of presentation. Following recall, the participant is asked a question about one detail from one sentence in the set. Each set become progressively longer by adding more sentences. The first set contained two sentences and the last set contained five sentences. This test was hand scored into scale scores.

The second task was the visual matrix subtest from the Swanson-Cognitive Processing Task (S-CPT) (Swanson, 1992). Children were shown
several dots within a matrix. The matrix was removed and the child was required to recall the number of dots within a specific column. Each matrix increased in the number of dots presented. The first matrix had two dots in a 2 x 2 matrix. The final matrix had twelve dots in a 9 x 5 matrix. The S-CPT has been recommended for people between the ages of five and eighteen (Swanson, 1996). The test was hand scored into scale scores.

The matrix span task is very similar to Ravens Standard Matrices Test, if only superficially. Both require identification of elements within a matrix. The matrix span task, however, requires recall of the elements of the matrix, a working memory task. The Raven’s Standard Matrices Test, on the other hand, requires reasoning across two dimensions in order to make a correct choice (Kane, Hambrick, Tuholski, Wilhelm, Payne & Engle, 2004). In the Kane et al study, a moderate correlation (0.42) was found between a matrix task and the Raven’s. The results of an exploratory factor analysis and structural equation modeling analysis show that the matrix span task loaded onto an identified working memory construct, while the Raven’s did not. This can be interpreted to suggest that reasoning tasks, like the Raven’s, involve working memory, but the two constructs and tasks are not identical.

Procedure

All participants were tested individually in a quiet room at their public school over two sessions. During the first session the Wisconsin Card Sorting
Task, the sentence span task, and the visual matrix task were administered. During the second session the math test, vocabulary test, and the Raven's Progressive Matrices were administered.

For all tasks, participants selected an answer and then gave a rating to represent how confident they were that they had identified the correct answer. For each item confidence ratings were presented in a Likert-scale format in the form of faces and percentages (Laupa, 1995). The five faces ranged from a very happy face to a very sad face. Under each face was a percentage representing the degree of confidence ranging from 100% (under the happiest face) to 0% (under the saddest face). The 50% mark had a neutral face. The scale was explained to each participant until it was clear that the child understood how to use it. Each participant had the option to circle either the face or the percentage to indicate his or her confidence for each item.
CHAPTER 4

RESULTS

There were two predictions in this study. The first prediction was that younger students would rely more heavily on domain-general metacognitive monitoring than domain-specific metacognitive monitoring; in other words, there would be moderate to strong positive correlations across all four tasks between performance and metacognitive monitoring scores with the youngest students. With older students, on the other hand, weak, positive or no significant correlations between the tasks would be observed. The second prediction was that older students would be more accurate when required to engage in domain-specific metacognitive monitoring than younger students. The next section explains the descriptive data, including the correlational data related to the first prediction. Then the multivariate data will then be presented in relationship to the second and third predictions.

Performance Scores

Raw scores from Raven’s Progressive Matrices Test, the WRAT Math, and WRAT Reading subtests were converted to the appropriate percentile ranking based on the specific norming tables for each test and grade. For the WCST, each participant’s standard score was determined by the percentage of errors
made during the test. The percentage of errors was then converted to a percentile ranking based on the norm tables. Means and standard deviations of all four tests for each grade can be found in Table 6. A preliminary analysis considered whether differences between males and females existed. No sex differences were observed.

Table 6. Means and Standard Deviations for Percentile Ranks listed by grade.

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th>Fourth</th>
<th>Seventh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Math</td>
<td>53.10</td>
<td>26.19</td>
<td>78.73</td>
</tr>
<tr>
<td>Reading</td>
<td>63.70</td>
<td>27.95</td>
<td>67.70</td>
</tr>
<tr>
<td>Raven's</td>
<td>55.67</td>
<td>27.22</td>
<td>65.13</td>
</tr>
<tr>
<td>WCST</td>
<td>39.33</td>
<td>34.50</td>
<td>68.00</td>
</tr>
</tbody>
</table>

Metacognitive Monitoring

The calibration accuracy quotient (CAQ) is an absolute index of calibration or a measure of discrimination (Nelson, 1996; Keleman et al., 2000). The CAQ reflects the degree to which a person’s confidence for a correct answer exceeds their confidence for an incorrect answer (Keren, 1991; Lundeberg, Fox, & Puncochar, 1994; Meeter & Nelson, 2003). A negative CAQ value represents higher confidence with wrong items compared to correct items. A positive CAQ value represents higher confidence with correct items and lower confidence with incorrect items. A CAQ of zero reflects an inability on the subject’s part to distinguish between right or wrong responses. The CAQ is the most common calibration index reported in educational research. All correct and incorrect
responses were used in tallying the CAQ scores including the ten consecutive error scores which determined the ending of testing in both the WRAT math and reading sections. CAQ scores could not be computed for 14 first grade students and two fourth grades students because of an absence of variance in confidence judgments (i.e., for example, a student consistently said they were 100% positive that they were right). The lowest CAQ score was -1.17, and the highest CAQ score was 4.47. CAQ means and standard deviations for the three grades can be found in Table 7.

Table 7. Means and standard deviations for CAQ scores listed by grade.

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th></th>
<th>Fourth</th>
<th></th>
<th>Seventh</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Math</td>
<td>2.30</td>
<td>1.18</td>
<td>1.47</td>
<td>0.52</td>
<td>1.03</td>
<td>0.64</td>
</tr>
<tr>
<td>Reading</td>
<td>1.41</td>
<td>0.78</td>
<td>1.43</td>
<td>0.31</td>
<td>1.45</td>
<td>0.36</td>
</tr>
<tr>
<td>Raven's</td>
<td>0.57</td>
<td>0.41</td>
<td>0.98</td>
<td>0.48</td>
<td>1.38</td>
<td>0.50</td>
</tr>
<tr>
<td>WCST</td>
<td>0.25</td>
<td>0.51</td>
<td>0.54</td>
<td>0.38</td>
<td>0.64</td>
<td>0.35</td>
</tr>
</tbody>
</table>

The Goodman-Kruskal gamma coefficient is considered the best relative indicator of metacognitive monitoring (Nelson, 1996; Kelemen, Frost & Weaver, 2000). Gamma coefficients indicate the accuracy of one item relative to other items regardless of measurement or judged magnitude (Meeter & Nelson, 2003). One strength of the gamma coefficient is the lack of susceptibility to unwanted influences, such as guessing (Nelson, 1996). Gamma coefficients consist of rank correlations and a range between 1 and -1 (Nelson, 1984). A positive score reflects high calibration ability, whereas, a negative score reflects low calibration.
ability. The Goodman-Kruskal gamma is the most common calibration index reported in psychological research. All correct and incorrect responses were used in tallying the gamma coefficients including the ten consecutive error scores which determined the ending of testing in both the WRAT math and reading sections. Gammas could not be determined for 14 first graders, two fourth graders, and two seventh graders in at least one task because these individuals used the same rating for both correct and incorrect responses. The means and standard deviations of gamma coefficients for the three grades are found in Table 8.

Table 8. Means and standard deviations for Goodman-Kruskal gamma coefficients identified by grade.

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th>Fourth</th>
<th>Seventh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Math</td>
<td>0.89</td>
<td>0.21</td>
<td>0.88</td>
</tr>
<tr>
<td>Reading</td>
<td>0.78</td>
<td>0.35</td>
<td>0.92</td>
</tr>
<tr>
<td>Raven's</td>
<td>0.68</td>
<td>0.29</td>
<td>0.75</td>
</tr>
<tr>
<td>WCST</td>
<td>0.21</td>
<td>0.38</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Bias scores measure the degree to which a participant was overconfident or underconfident in relationship to item accuracy. The bias score is the difference between the mean level of confidence and the mean performance score divided by 100. Bias scores range from -1 to +1. Scores greater than zero represent overconfidence while scores less than zero represent underconfidence. A score close to zero represents no bias or an accurate judgment. All correct and
incorrect responses were used in tallying the bias scores including the ten consecutive error scores which determined the ending of testing in both the WRAT math and reading sections. The means and standard deviations for bias scores are in Table 9.

Table 9. Means and standard deviations for bias scores listed by grade.

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th>Fourth</th>
<th>Seventh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Math</td>
<td>0.69</td>
<td>0.16</td>
<td>0.51</td>
</tr>
<tr>
<td>Reading</td>
<td>0.61</td>
<td>0.13</td>
<td>0.40</td>
</tr>
<tr>
<td>Raven's</td>
<td>0.67</td>
<td>0.20</td>
<td>0.49</td>
</tr>
<tr>
<td>WCST</td>
<td>0.11</td>
<td>0.24</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Correlations

Math and reading percentile rankings were positively correlated for first grade (r = 0.64) and fourth grade (r = .57) but not for seventh grade students (see Tables 10 through 12 respectively). Math percentile rankings were positively correlated with Raven’s Progressive Matrices percentile rankings for the fourth (r = .57) and seventh (r = .65) grade students, but not for the first graders. No significant correlations between math percentile rankings and WCST percentile rankings were observed. WRAT math percentile rankings were positively correlated with scores on the Raven’s for fourth grade students only (r = .56). WCST percentile rankings and percentile rankings on the Raven’s were positively correlated for first grade students only (r = .48).
Table 10. Correlations of percentile rank for first grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven's</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.64*</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>0.22</td>
<td>0.34</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>0.21</td>
<td>0.10</td>
<td>0.48*</td>
<td>1.00</td>
</tr>
</tbody>
</table>
* = p < 0.01

Table 11. Correlations of percentile rank for fourth grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven's</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.57*</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>0.57*</td>
<td>0.56*</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>0.11</td>
<td>0.06</td>
<td>0.19</td>
<td>1.00</td>
</tr>
</tbody>
</table>
* p < 0.01

Table 12. Correlations of percentile rank for seventh grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven's</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.27</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>0.65*</td>
<td>0.35</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>0.28</td>
<td>0.26</td>
<td>0.25</td>
<td>1.00</td>
</tr>
</tbody>
</table>
* p < 0.01

CAQ scores were not significantly correlated with each other for any of the performance measures for first or seventh grade students (see Tables 13 and 15). Three moderate, positive correlations were observed for the fourth grade students (see Table 14). Significant positive relationships were observed between the WCST and the WRAT math scale (r = 0.38), between Raven's and
WRAT reading scale \( (r = 0.52) \), and between the WRAT math and WRAT reading scales \( (r = .59) \).

Table 13. Correlations of CAQ scores for first grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven's</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.38</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>-0.02</td>
<td>0.31</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>0.26</td>
<td>0.18</td>
<td>0.09</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 14. Correlations of CAQ scores for fourth grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven's</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.59**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>0.38*</td>
<td>0.52**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>0.39*</td>
<td>0.31</td>
<td>0.19</td>
<td>1.00</td>
</tr>
</tbody>
</table>
* = \( p < 0.05 \)
** = \( p < 0.01 \)

Table 15. Correlations of CAQ scores for seventh grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven's</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>-0.17</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>0.03</td>
<td>0.26</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>-0.09</td>
<td>0.22</td>
<td>0.33</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Gamma coefficients were not significantly correlated between tasks for either first grade students (see Table 16) or seventh grade students (see Table 18). One moderate, positive correlation among the fourth grade students occurred between the Raven's and the WRAT reading scale \( (r = .38) \) (see Table 17).
Table 16. Correlations of Goodman-Kruskal gamma coefficients for first grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven's</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.27</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>0.23</td>
<td>-0.18</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>0.00</td>
<td>0.28</td>
<td>-0.16</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 17. Correlations of Goodman-Kruskal gamma coefficients for fourth grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven's</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.22</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>0.03</td>
<td>0.38*</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>0.29</td>
<td>0.15</td>
<td>0.03</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* = p < 0.05

Table 18. Correlations of Goodman-Kruskal gamma coefficients for seventh grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven's</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>-0.03</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>-0.15</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>-0.10</td>
<td>-0.20</td>
<td>0.07</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Several bias scores were significantly correlated between tasks for first grade students (see Table 19): the WCST and Raven's (r = -0.51), the WCST and the WRAT reading scale (r = 0.64), and Raven’s and the WRAT reading scale (r = 0.64). Significant correlations were observed for fourth grade students between WRAT math and WRAT reading (r = .38); WRAT math and Raven’s (r =
.41); WRAT reading and WCST (r = .36); and Raven’s and WCST (r = .51) (see Table 20). No significant correlations between bias scores were observed for the seventh grade students (see Table 21).

Table 19. Correlations of bias scores for first grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.19</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>-0.06</td>
<td>0.64*</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>-0.12</td>
<td>0.64*</td>
<td>-0.51*</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* = p < 0.01

Table 20. Correlations of bias scores for fourth grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.38*</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>0.41*</td>
<td>0.31</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>0.34</td>
<td>0.36*</td>
<td>0.51**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* = p < 0.05
** = p < 0.01

Table 21. Correlations of bias scores for seventh grade students.

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Reading</th>
<th>Raven</th>
<th>WCST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.16</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven</td>
<td>-0.19</td>
<td>-0.18</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>-0.15</td>
<td>0.27</td>
<td>0.16</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* = p < 0.01

To further investigate the first hypothesis a repeated measures analysis of variance (ANOVA) for each grade was conducted. It was expected that monitoring scores would be equivalent among first grade students across all
tasks, but older students should have higher monitoring scores in the domain-specific tasks, than the domain-general tasks. If this was the result it would suggest domain-general monitoring was used as a default strategy until sufficient practice had occurred in the reading and mathematic content areas. CAQ scores on all four tasks, WRAT math, WRAT reading, the Raven’s, and the WCST served as dependent variables. A significant effect was observed for the first grade students, \( F(2,28) = 30.74, p < 0.05, \eta^2 = 0.67 \) (the assumption of sphericity was violated, so the Greenhouse-Geisser correction was employed, Keppel, 1991). Follow-up Tukey HSD analyses indicated that the first grade students were less accurate at absolute metacognitive monitoring on the WCST (\( M = 0.25 \)) than on the WRAT math (\( M = 2.30 \)) and the WRAT reading tests (\( M = 1.42 \)), but not the Raven’s (\( M = 0.57 \)). The first grade students were also less confident on the Raven’s than on both the WRAT math and WRAT reading tests. In addition, the first grade students were less confident on the WRAT math than on the WRAT reading test.

For the fourth grade students, CAQ scores on all four tasks, WRAT math, WRAT reading, the Raven’s and the WCST served as dependent variables. The assumption of sphericity was upheld. A significant effect was observed for the fourth grade students, \( F(3,81) = 45.13, p < 0.05, \eta^2 = 0.63 \). Follow-up Tukey HSD analysis indicated that fourth grade students were less accurate at absolute metacognitive monitoring on the WCST (\( M = 0.54 \)) than on the Raven’s (\( M = 0.98 \)), the WRAT math (\( M = 1.47 \)), and the WRAT reading tests (\( M = 1.43 \)).
Fourth grade students were also less confident on the Raven’s than the WRAT math and WRAT reading tests.

For the seventh grade students’, CAQ scores on all four tasks, WRAT math, WRAT reading, the Raven’s and the WCST served as dependent variables. A significant effect was observed for the seventh grade students, $F(2,61) = 18.88$, $p < 0.05$, $\eta^2 = 0.40$ (the assumption of sphericity was violated, so the Greenhouse-Geisser correction was employed, Keppel, 1991). Follow-up Tukey HSD analysis indicated that seventh grade students were less accurate at absolute metacognitive monitoring on the WCST ($M = 0.64$) than from Ravens ($M = 1.38$), the WRAT math ($M = 1.03$), and the WRAT reading tests ($M = 1.45$). Seventh grade students were more accurate at metacognitive monitoring on the WRAT math than the Raven’s and the WRAT reading test.

For the first grade gamma coefficients on all four tasks, WRAT math, WRAT reading, the Raven’s, and the WCST served as dependent variables. The assumption of sphericity was upheld. A significant effect was observed for the first grade students, $F(3,45) = 18.45$, $p < 0.05$, $\eta^2 = 0.55$. Follow-up Tukey HSD analysis indicated that first grade students less accurate at relative metacognitive monitoring on the WCST ($M = 0.21$) than the Raven’s ($M = 0.68$), WRAT math ($M = 0.88$), and WRAT reading tests ($M = 0.78$).

The fourth grade students’ gamma coefficients on all four tasks, WRAT math, WRAT reading, the Raven’s, and the WCST served as dependent variables. A significant effect was observed for the fourth grade students, $F(2,57) = 55.42$, $p$
< 0.05, η² = 0.67 (the assumption of sphericity was violated, so the Greenhouse-Geisser correction was employed, Keppel, 1991). Follow-up Tukey HSD analysis indicated that the fourth grade students were less accurate at relative metacognitive monitoring on the WCST (M = 0.38) than the Raven's (M = 0.75), the WRAT math (M = 0.88), and the WRAT reading tests (M = 0.92). Fourth grade students were also less accurate at monitoring the Raven’s than the WRAT math and WRAT reading test.

The seventh grade students' gamma coefficients on all four tasks, WRAT math, WRAT reading, the Raven's and the WCST, served as dependent variables. A significant effect was observed for the seventh grade students, F(2,51) = 14.46, p < 0.05, η² = 0.34 (the assumption of sphericity was violated, so the Greenhouse-Geisser correction was employed, Keppel, 1991). Follow-up Tukey HSD analysis indicated that the seventh grade students were less accurate at relative metacognitive monitoring on the WCST (M = 0.47) than for the Raven’s (M = 0.85), the WRAT math (M = 0.68), and the WRAT reading tests (M = 0.89). Seventh grade students were also less accurate at monitoring the WRAT math than the WRAT reading test.

The first grade students' bias scores on all four tasks, WRAT math, WRAT reading, the Raven's, and the WCST served as dependent variables. A significant effect was observed for the first grade students, F(2,58) = 89.07, p < 0.05, η² = 0.75 (the assumption of sphericity was violated, so the Greenhouse-Geisser correction was employed, Keppel, 1991). Follow-up Tukey HSD analysis indicated
that first grade students showed less overconfidence on the WCST (M = 0.11) than on the Raven’s (M = 0.67), the WRAT math (M = 0.69), and the WRAT reading tests (M = 0.61).

The fourth grade students’ bias scores on all four tasks, WRAT math, WRAT reading, the Raven’s and the WCST, served as dependent variables. The assumption of sphericity was upheld. A significant effect was observed for the fourth grade students, $F(2,87) = 177.18, p < 0.05, \eta^2 = 0.86$. Follow-up Tukey HSD analysis indicated that fourth grade students were less overconfident on the WCST (M = -0.01) than on the Raven’s (M = 0.49), the WRAT math (M = 0.51), and the WRAT reading tests (M = 0.40). Fourth grade students showed less overconfidence on the WRAT reading test than on the Raven’s and the WRAT math test.

The seventh grade students’ bias scores on all four tasks, WRAT math, WRAT reading, the Raven’s, and the WCST served as dependent variables. The assumption of sphericity was upheld. A significant effect was observed for the seventh grade students, $F(3,90) = 75.09, p < 0.05, \eta^2 = 0.72$. Follow-up Tukey HSD analysis indicated that seventh grade students were less overconfident on the WCST (M = -0.04) than on the Raven’s (M = 0.43), the WRAT math (M = 0.45), and the WRAT reading tests (M = 0.32). Seventh grade students were also less overconfident the on WRAT reading test than on the Raven’s and the WRAT math test.

*Principal Component Analysis*
Principal component analysis was conducted on all percentile rankings in order to determine the underlying structure of performance percentile rankings, CAQ scores, gamma coefficients, and bias scores to validate whether the four tasks could be classified as either domain-specific or domain-general tasks prior to further analysis. Principal component analysis would indicate whether metacognitive monitoring could be considered predominantly domain-general, domain-specific, or some combination. The presence of only one component would suggest that metacognitive monitoring is predominantly domain-general; whereas the presence of four components would suggest metacognitive monitoring is predominantly domain-specific. It was expected that two components would appear: the first consisting of the WCST and the Raven’s representing domain-general metacognitive monitoring and the second consisting of the two WRAT subscales representing domain-specific metacognitive monitoring. Components were determined from an examination of eigenvalues and analysis of the scree plot.

The initial percentile rank analysis was for performance scores. One component resulted (with varimax rotation) that accounted for 51.84% of the variance. Factor loadings are listed in Table 22. A second factor analysis was conducted in order to examine the prediction that the four tests would separate into two constructs, one representing the domain-specific tests and one representing the domain-general tests. The factor loadings from this analysis are pictured in Table 23. The first component consisted of the WRAT math and
reading scales and Raven's. The second component contained the WCST. This model accounted for 73.29% of the variance.

Table 22. Principal component analysis loadings for percentile rankings across all three grades.

<table>
<thead>
<tr>
<th></th>
<th>Component 1</th>
<th></th>
<th>Component 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven's</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WCST</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 23. Two component model component loadings for percentile rankings

<table>
<thead>
<tr>
<th></th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>0.76</td>
<td>0.27</td>
</tr>
<tr>
<td>Reading</td>
<td>0.85</td>
<td>-0.09</td>
</tr>
<tr>
<td>Raven's</td>
<td>0.71</td>
<td>0.36</td>
</tr>
<tr>
<td>WCST</td>
<td>0.13</td>
<td>0.96</td>
</tr>
</tbody>
</table>

A principal component analysis, with varimax rotation, was conducted on the CAQ scores. Two components were extracted based on an examination of the eigenvalues and the scree plot. The first component consisted of the WRAT math task and the WRAT reading tasks (see Table 24). The second component consisted of Raven's. The WCST loaded onto both components. These two components accounted for 71.88% of the variance.

Two components were also derived from a principal component analysis of the gamma coefficients. The first component consisted of the WRAT math and reading tasks as well as the WCST (see Table 25). The second component...
Table 24. Two component model component loadings for CAQ scores.

<table>
<thead>
<tr>
<th>Test</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>0.73</td>
<td>0.07</td>
</tr>
<tr>
<td>Reading</td>
<td>0.78</td>
<td>-0.31</td>
</tr>
<tr>
<td>Raven's</td>
<td>-0.10</td>
<td>0.92</td>
</tr>
<tr>
<td>WCST</td>
<td>0.66</td>
<td>0.56</td>
</tr>
</tbody>
</table>

included only Raven’s. These two components accounted for 62.88% of the variance.

Table 25. Two component model component loadings for gamma coefficients

<table>
<thead>
<tr>
<th>Test</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>0.49</td>
<td>-0.35</td>
</tr>
<tr>
<td>Reading</td>
<td>0.79</td>
<td>-0.08</td>
</tr>
<tr>
<td>Raven’s</td>
<td>0.04</td>
<td>0.92</td>
</tr>
<tr>
<td>WCST</td>
<td>0.77</td>
<td>0.27</td>
</tr>
</tbody>
</table>

The initial principal component analysis for bias scores produced one component that accounted for 58.32% of the variance (see factor loadings in Table 26). A second factor analysis was conducted in order to examine the prediction that the four tests would separate into two constructs, one representing the domain-specific tests and one representing the domain-general tests. The factor loadings from this analysis are pictured in Table 27. In this two-factor solution, the first component consisted of the domain-specific tasks: WRAT math and WRAT reading, as well as the WCST. The second component consisted of the domain-general tasks, the WCST and Raven’s. The two components accounted for 79.59% of the variance.
Table 26. One component model component loadings for bias scores

<table>
<thead>
<tr>
<th>Component 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
</tr>
<tr>
<td>Reading</td>
</tr>
<tr>
<td>Raven’s</td>
</tr>
<tr>
<td>WCST</td>
</tr>
</tbody>
</table>

Table 27. Two component model component loadings for bias scores

<table>
<thead>
<tr>
<th></th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>0.90</td>
<td>0.05</td>
</tr>
<tr>
<td>Reading</td>
<td>0.77</td>
<td>0.28</td>
</tr>
<tr>
<td>Raven’s</td>
<td>0.08</td>
<td>0.96</td>
</tr>
<tr>
<td>WCST</td>
<td>0.58</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Two components were anticipated due to the results of previous research. Similar results have been obtained in previous research. Schraw and Nietfeld (1998; Schraw, Dunkle, Bendixen & Roedel, 1995; Veenman & Verhij, 2003), for example, described a two-factor model with three domain-general metacognitive monitoring tasks loading onto one component and five domain-specific tasks loading onto a second component. While there was some consistency between this research and previous studies in that two components were derived, the WCST did not load on the domain-general metacognitive monitoring component as anticipated. Based on the results of the factor analysis, it was decided that three variables would be used in the multivariate analysis. Ravens and WCST would be considered as separate variables, while a composite score of the WRAT math and WRAT reading subscales would represent the domain-specific metacognitive monitoring variable. Using a composite score to represent domain-
specific monitoring was used to simplify the analysis between domain-specific monitoring and the domain-general tasks. The composite score was the average of each subject's WRAT math and WRAT reading score.

Correlations between Raven's and the domain-specific composite scores were computed across all three grades. Domain-specific composite percentile rankings correlated significantly with the Raven's ($r = 0.52$) percentile rankings. Domain-specific monitoring CAQ scores did not correlate significantly with the Raven's ($r = -0.06$). Domain-specific composite gamma coefficients did not correlate significantly with the Raven's ($r = 0.04$). Domain-specific monitoring bias scores correlated significantly with the Raven's ($r = 0.48$).

**Multivariate Analysis**

For the performance percentile rankings, a one-way multivariate analysis of covariance (MANCOVA) was conducted. The independent variable was grade level (first, fourth, seventh). Percentile ranking scores on the WRAT math and WRAT reading were condensed into a composite score that represented domain-specific metacognitive monitoring. Percentile rankings on the Raven's and the WCST served as the other dependent variables. Performance on sentence span and matrix span served as covariates. Neither sentence span nor matrix span served as significant covariates, so both were dropped from the analysis. A one-way multivariate analysis of variance (MANOVA) was then conducted with the independent and dependent variables described above. A significant effect for
grade level was found; Wilks $\lambda_i = 0.79$, $F_{(6,172)} = 3.64$, $p < 0.05$, $\eta^2 = 0.11$.

Follow up univariate analyses of variance indicated significant differences between grade levels on the WCST and the domain-specific monitoring composite score. Post hoc Tukey HSD on the WCST showed that first grade students ($M = 39.33$) had significantly lower percentile rank scores than fourth ($M = 68.00$) and seventh ($M = 62.84$) grade students. Post hoc Tukey HSD on the domain-specific composite percentile ranking showed that first grade students ($M = 58.40$) had significantly lower percentile rankings than fourth grade students ($M = 73.22$).

For the CAQ scores a one-way MANCOVA was conducted. The independent variable was grade level (first, fourth, seventh). CAQ scores on the WRAT math and reading condensed into a composite dependent variable. CAQ scores on the Raven's and the WCST served as the other two dependent variables. Performance on sentence span and matrix span served as covariates. Matrix span did not significantly contribute to the analysis, but sentence span did. The matrix span task was dropped from the next analysis. A one-way MANCOVA with sentence span as the single covariate was conducted. The sentence span covariate was significant; Wilks $\lambda_i = 0.85$, $F_{(3,67)} = 3.90$, $p < 0.05$, $\eta^2 = 0.15$. A significant effect for grade level was found; Wilks $\lambda_i = 0.58$, $F_{(6,134)} = 6.97$, $p < 0.05$, $\eta^2 = 0.24$. Follow-up univariate analysis indicated significant differences between grade levels on Raven's and the domain-specific monitoring composite score. Post hoc Tukey HSD on Raven's showed seventh grade
students (M = 1.38) had higher CAQ scores than first (M = 0.57) and fourth grade students (M = 0.98). Post hoc Tukey on the domain-specific metacognitive monitoring composite CAQ score showed that first grade students (M = 1.90) had higher CAQ scores than fourth (M = 1.46) and seventh grade students (M = 1.22).

For the gamma coefficients a one-way MANCOVA was conducted. The independent variable remained the three grade levels and the three dependent variables remained as the WCST, Raven's, and the domain-specific metacognitive monitoring composite. Performance on sentence span and matrix span served as covariates. Matrix span did not significantly contribute to the analysis, but sentence span did. The matrix span task was dropped from the next analysis. A one-way MANCOVA with sentence span as the single covariate was conducted. The sentence span covariate was significant; Wilks $\lambda = 0.89$, $F_{(3,67)} = 2.79$, $p < 0.05$, $\eta^2 = 0.11$. No significant effect for grade level was found; Wilks $\lambda = 0.83$, $F_{(6,134)} = 2.13$, $p = 0.054$, $\eta^2 = 0.09$.

For the bias scores a one-way MANCOVA was conducted. The independent variable remained the three grade levels and the three dependent variables remained as the WCST, Raven's, and the domain-specific metacognitive monitoring composite. Performance on sentence span and matrix span served as covariates. Neither sentence span nor matrix span served as significant covariates, so both were dropped from the analysis. A one-way MANOVA was then conducted with the independent and dependent variables described above.
A significant effect for grade level was found; Wilks $\lambda = 0.40$, $F_{(5,134)} = 16.43$, $p < 0.05$, $\eta^2 = 0.36)$. Follow-up univariate analysis indicated significant differences between grade levels on the WCST, Raven's and the domain-specific monitoring composite score. Post hoc Tukey HSD on the WCST showed that first grade students ($M = 0.11$) had significantly higher bias scores than fourth ($M = -0.01$) and seventh ($M = -0.04$) grade students. Post hoc Tukey HSD on Raven's showed first grade students ($M = 0.67$) had higher bias scores than fourth ($M = 0.49$) and seventh grade students ($M = 0.43$). Post hoc Tukey on the domain-specific monitoring composite bias score showed that first grade students ($M = 0.65$) had higher bias scores than fourth ($M = 0.45$) and seventh grade students ($M = 0.38$). Fourth grade student bias scores were significantly higher than seventh grade students.
CHAPTER 5

DISCUSSION

The purpose of this study was to identify the developmental trend of domain-general and domain-specific metacognitive monitoring in children. There were two research questions. The first question was which appears first domain-general metacognitive monitoring or domain-specific metacognitive monitoring? Currently no research studies have documented whether domain-general metacognitive monitoring or domain-specific metacognitive monitoring appears first.

The second question was what are the differences between domain-general metacognitive monitoring and domain-specific metacognitive monitoring across a specific developmental timeline? Current theory suggests that domain-general metacognitive monitoring improves as a function of the biological maturation of working memory and less as a response to experience. Unfortunately, there is limited research in this area. The research that does exist tends to support the hypothesis that domain-specific metacognitive monitoring improves in a linear fashion as a result of practice and instruction (Bisanz et al., 1978; Short et al., 1993), and is less dependent on the development of working memory (Bayliss, Jarrold, Gunn, & Baddeley, 2003). The developmental progression of domain-general metacognitive monitoring and domain-specific
metacognitive monitoring had not been explicitly examined prior to this study. The present study investigated whether domain-general and domain-specific metacognitive monitoring are present in first grade children. It also attempted to track, through a cross sectional research approach, the progression of each type of metacognitive monitoring from first to fourth to seventh grade.

Two competing hypotheses were considered. The experiential hypothesis, which is based on the research of Schraw and Nietfeld (1998), would require that domain-specific metacognitive monitoring appear first with domain-general metacognitive monitoring appearing later. The modularization hypothesis, on the other hand, would require that domain-general metacognitive monitoring appear first with domain-specific metacognitive monitoring appearing later (e.g. Karmiloff-Smith, 1992; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Karmiloff-Smith, Plunkett, Johnson, Elman, & Bates, 1998). Each hypothesis would support the improvement of both domain-general metacognitive monitoring and domain-specific metacognitive monitoring with experience.

If the modularization hypothesis is true, domain-general metacognitive monitoring is the default process and domain-specific metacognitive monitoring appears after substantial experience within a specific content domain. Domain-general metacognitive monitoring, therefore, would improve over time, primarily as a function of biological maturation, while domain-specific metacognitive monitoring would improve as a result of experience.
If the experiential hypothesis is correct, both domain-general and domain-specific metacognitive monitoring would improve as a result of experience. Domain-general monitoring would appear in the later grades as a function of transfer of metacognitive and strategy knowledge across content domains.

These hypotheses were tested using a number of measures including the Raven’s Standard Progressive Matrices Test, the Wisconsin Card Sorting Task, the Wide Range Achievement math subtest, and the Wide Range Achievement reading subtest. The first prediction was that domain-general metacognitive monitoring would appear first followed by domain-specific metacognitive monitoring. In addition, it was predicted that only domain-specific metacognitive monitoring would improve with experience. In other words, the prediction was that the data would support the modularization hypothesis and not the experiential hypothesis.

Review of Results

The results from this study can be summed up as follows: strong, positive correlations occurred predominantly with the fourth grade students. There were no significant correlations between tasks among seventh grade students. Working memory, in the form of a sentence span task, had an influence on domain-general and domain-specific metacognitive monitoring accuracy, but did not affect achievement or overconfidence. Older students were more accurate in performance and monitoring than younger students on the domain-general
metacognitive monitoring tasks. The results provide support for the experiential hypothesis.

There were moderate positive correlations between tasks with first grade students for the performance percentile rankings and bias scores, but weak or no correlations between tasks for the CAQ scores or the gamma coefficients. There were moderate positive correlations across all tasks for the fourth grade students. There were weak correlations between tasks among the seventh grade students. The results of the principal component analysis support the presence of one domain-general metacognitive monitoring process. The pattern of correlations between tasks from grade to grade parallels the work of Schraw and colleagues (Schraw, Dunkle, Bendixen et al., 1995; Schraw & Graham, 1997; Schraw & Nietfield, 1998) who also found moderate correlations between tasks in bias scores and monitoring accuracy. One difference between the results of this study and Schraw’s work is the absence of significant correlations between tasks for the seventh grade students. These results may be interpreted as support for greater reliance on domain-specific metacognitive processes in seventh grade students.

The analysis of the CAQ scores indicated that first grade students were monitoring most accurately on the WRAT math, followed by the WRAT reading, the Raven’s, and finally the WCST. Fourth grade students monitored equally well on the WRAT math and WRAT reading tasks. Fourth grade students were less accurate monitors on the Raven’s compared to the domain-specific tasks and the
least accurate at monitoring the WCST. Seventh grade students were equally accurate at monitoring the WRAT math, WRAT reading, and the Raven’s. Seventh grade students were less accurate at monitoring the WCST.

The analysis of the gamma coefficients indicated that first grade students were monitoring most accurately on the WRAT math and WRAT reading tasks. The first grade students were more accurate when monitoring the Raven’s than the WCST. Fourth grade students were more accurate at monitoring the WRAT math and WRAT reading tasks than either the Raven’s or the WCST. Fourth grade students were more accurate at monitoring the Raven’s than the WCST. Seventh grade students were most accurate at the WRAT reading, followed by the Raven’s, then the WRAT math, and finally the WCST.

The analysis of the bias scores indicated that first grade students were more overconfident on the WRAT math, WRAT reading, and the Raven’s tasks than the WCST. Fourth grade students were more overconfident on WRAT math and WRAT reading than the Raven’s and the WCST. Fourth grade students showed greater overconfidence on the Raven’s when compared to the WCST. Seventh grade students were more overconfident on WRAT reading and the Raven’s than WRAT math and the WCST. Seventh grade students showed greater overconfidence on the WRAT math when compared to the WCST.

One apparent inconsistency was between the high CAQ scores among first grade students on the domain-specific tasks but high overconfidence as measured by bias scores on the same tasks. This may be interpreted as follows;

92
the first grade students demonstrated good discrimination between correct and incorrect responses although they were overall overconfident in their ability to monitor their performance. Contrast these results with the seventh grade students' performance on the Raven's. These students demonstrated good discrimination between correct and incorrect responses, as measure by CAQ scores, and low overconfidence as measured by bias scores. The seventh grade students were better overall at monitoring than first grade students, who can discriminate what content knowledge they have from the content knowledge they lack.

Performance, as measured by percentile ranks, was significantly different across grades, which would be expected. Overall, first grade students had lower performance scores than the fourth and seventh grade students on the WCST and the domain-specific metacognitive monitoring composite (math and reading) scores. Fourth grade students had higher performance scores on the Raven's than first and seventh grade students.

With respect to calibration accuracy (CAQ scores), seventh grade students demonstrated greater absolute monitoring accuracy for the Raven's while first grade students had greater absolute monitoring accuracy on the domain-specific composite (the WRAT math and the WRAT reading) (see Figure 3). There were no differences between grades and tasks as measured by gamma coefficients (see Figure 4).

With respect to bias scores, first grade students demonstrated greater
overconfidence than fourth or seventh grade students on the WCST, the Raven's, and the domain-specific monitoring composite.

The results of this study may be interpreted as support for the hypothesis that domain-specific metacognitive monitoring occurs first: the experiential hypothesis. In addition, the data provides evidence that domain-specific metacognitive monitoring improves with practice with respect to absolute accuracy. These results fit nicely with the work of Myers and Paris (1978) who suggested that reading monitoring, a domain-specific activity, requires both maturation and skill development. In the current study, improvement in metacognitive monitoring accuracy, as determined by the measure of
overconfidence (bias scores) from first to fourth to seventh grade was observed for all tasks. Students from all three grades, however, were more accurate when monitoring the Ravens’ than when monitoring the WCST.

**Explanation of Results**

The absence of significant correlations for first grade students between the WCST, Raven’s, the WRAT math and WRAT reading tests on the CAQ and the gamma coefficients may be interpreted as support for domain-specific metacognitive monitoring appearing first, which would support the experiential hypothesis. The absence of significant correlations for seventh grade students between the WCST, Raven’s, the WRAT math and WRAT reading composite on percentile rankings, the CAQ scores, the gamma coefficients, and the bias scores may support the claim that domain-general metacognitive monitoring appears
later in the developmental sequence. The multivariate analyses also provided support for the experiential hypothesis. Metacognitive monitoring, as measured by the CAQ scores, gamma coefficients, and bias scores, improved for the domain-general tasks (the WCST and Raven's) in that fourth grade students were more accurate and less overconfident than first grade students, and that seventh grade students were more accurate and less overconfident than fourth grade students. Metacognitive monitoring of the domain-specific tasks (the WRAT math and WRAT reading) improved for the bias scores in that fourth grade students were less overconfident than first grade students, and seventh grade students were less overconfident than fourth grade students.

Why does domain-general metacognitive knowledge appear to be influenced more by experience than maturation? One possible explanation is related to self-generated feedback. Flavell (1979) proposed that metacognitive monitoring improves as a result of internal feedback. Much metacognitive knowledge is obtained through experience. Internal feedback may not be limited to the domain or task that is being monitored but is applied to all metacognitive tasks that are monitored.

Another reason that metacognitive monitoring likely improves due to experience rather than through maturation is related to strategy instruction. Some authors have suggested reading instruction contains quite a bit of metacognition instruction (Brown et al., 1984; Delclos & Harrington, 1991; Paris & Jacobs, 1984). These researchers consider reading as a domain-general task.
(Brown & Palinscar, 1984). Brown and Palinscar (1984) suggest metacognitive monitoring of reading affects domain-specific knowledge since most content specific knowledge is gained through reading. On the other hand, several studies have documented that little metacognition instruction occurs with grade school children (Clift, et al, 1990). Further research is necessary to explain these findings.

An unexpected effect was the drop in monitoring accuracy from the first to the fourth and seventh grade students with respect to the domain-specific composite score, seen specifically in the WRAT math monitoring values. One explanation for the decrease in monitoring accuracy across grades with the domain-specific composite score may be that the older students move to a heuristic strategy in the domain-specific tasks, while maintaining reliance on simple metacognitive monitoring for the domain-general tasks. Current research suggests that when learners are attempting to recall specific information they tend to rely on cues from the prompt. Koriat's accessibility model (Koriat, 1993), for example, promotes this view. Students rely on heuristics instead of monitoring because the declarative knowledge is in long-term memory and can simply be retrieved. Heuristics include reliance on familiarity with the domain (Glenberg, Wilkenson, & Epstein, 1982; Maki, 1999), cue familiarity (Metcalfe, 1993; Miner & Reder, 1994), semantic attributes (Koriat, 1994), ease of retrieval (Koriat, 1998), and fluency of processing (Koriat, Bjork, Sheffer, & Bar, 2004). The nature of the domain-general tasks (the Raven's and the WCST) required
monitoring a problem while determining the solution that did not require specific content knowledge; rather, these the domain-general tasks rely more on reasoning. Further research would be necessary to validate this explanation.

Limitations of Present Study

The major limitation of this study was related to the WCST. The WCST, sample size and counterbalancing will be addressed in this section as limitations. The WCST did not function as was expected. In particular, in the factor analyses the WCST did not load with the Raven’s, which had been anticipated. It is possible it did not load as expected because it may be a measure of inhibition (Dempster & Corkill, 1999; Miyake, Friedman & Emerson, 2000; Andres, 2003) rather than a measure of domain-general ability (Chelune & Baer, 1986; Fristoe, Salthouse & Woodard, 2005; Laws, 1999). It is also possible it did not load as expected because of the set up of the test. In order to complete the task, examinees receive immediate feedback on a per item basis, whereas, no feedback, immediate or otherwise, was provided on the other three tasks. Use of a different measure might have been helpful. Unfortunately, the only domain-general ability task available that is appropriate for use with children in Kindergarten and/or first grade appears to be the Raven’s. The absence of alternative domain-general ability measures for such young children will likely make it difficult to more precisely understand the relationship between domain-specific and domain-general metacognitive monitoring for this age group.
Another limitation was the small number of students recruited from each grade. The results of the factor analyses may be unreliable due to the small number of subjects. Increasing the number of subjects threefold would have increased the reliability of the factor analysis. Given that this study was the first to consider the issue of metacognitive monitoring development for both domain-specific and domain-general a relatively small sample was appropriate.

Almost half of the first grade students’ CAQ scores and gamma coefficients had to be excluded from the analyses because those students selected the same confidence rating for every item within the task. When a student limits their responses, or in other words does not make any discrimination between correct and incorrect answers, neither a CAQ score nor a gamma coefficient can be computed. This inability on the part of the first grade students to distinguish between right and wrong answers should be examined in future research. One possible implication is that age of on-set for metacognitive monitoring is highly variable and begins somewhere between age 4 and forth grade.

Another limitation is related to the sequence of test presentations. All tasks were presented in the same order throughout the study due, in large part, to restrictions imposed by school district personnel. Teachers and building principals were concerned that children who participated be gone from their classroom for as short a time period as possible. The best response to this concern/restriction was to present the tests in a specific sequence in order to
minimize time spent with each student. Counterbalancing tasks would have been the preferred option, but was not possible given the administrative constraints.

The final limitation relates to the use of a cross-sectional quasi-experimental design. Cross-sectional designs allow for age difference comparisons; however, these differences cannot be attributed specifically to development (Sigelman, 1999). Furthermore, differential patterns of development cannot be identified. As a quasi-experimental design alternative explanations are more difficult to rule out (Shadish, Cook & Campbell, 2002).

*Future Directions*

While this study provided preliminary support for the role of experience in domain-general and domain-specific metacognitive monitoring, further research is needed. One issue that should be addressed is the absence of measures of domain-general ability for young children. Other researchers (e.g., Schraw & Nietfeld, 1998) have used the Schaie-Thurstone Letter Series for adults as a measure of domain-general ability; however, no equivalent is available for children. Other measures of domain-general ability exist for young children; they do not, however, lend themselves to a study of this nature because of how they are administered. The block design subtest from the Wechsler Intelligence Scale for Children, for example, is strongly, positively correlated with performance on the Raven’s (Martin & Wiechers, 1954). It would be just a problematic as the WCST, though, because there is little ambiguity in terms of whether the child has
provided the correct solution. That is, the child recreates a visible pattern—hence it is abundantly clear to the child whether they have accurately done so. Furthermore, the block design is scored on speed. The administration and scoring procedures, therefore, prohibit the inclusion of confidence judgments that would be required. There simply is no readily available instrument (D. Allen, personal communication, February 21, 2007). The development of a new instrument that has a similar format to the Raven's that is appropriate for use with small children would be extremely helpful.

Another area for consideration is the role of strategy use/instruction. Research is clear that while strategy use can be spontaneous (Brown & Smiley, 1978), spontaneous strategy transfer is a different issue (Carr & Alexander, 1996; Fisher, 1998; Garner & Alexander, 1989; Kuhn, 1995; O'Sullivan & Pressley, 1984). To examine the absence of growth in metacognitive monitoring with respect to the reading and math scores, one study attempt to confirm the use of metacognitive monitoring strategies in reading and math. If reading comprehension strategies and mathematic monitoring strategies are not being taught would instruction in reading and math strategies be used in fourth and seventh grade students, or would these students rely more on heuristic strategies?
Main Contribution

The main contribution of this dissertation study is that it provides evidence that domain-specific metacognitive monitoring appears before domain-general metacognitive monitoring. Furthermore, this study provides evidence that both domain-general monitoring and domain-specific monitoring benefit from experience. Although the sentence span task, which was used as a covariate in an attempt to control for changes in working memory span from grade to grade (a quasi-maturational process) was significant in several analyses, maturational processes are likely less influential than previously believed. While the results of several studies suggested support for the experiential hypothesis, this was the first study to directly test both the experiential and modularization hypotheses. This study provides a foundation in guiding future research and conceptualizing the progression of domain-general and domain-specific metacognitive monitoring.
REFERENCES


Reference


VITA

Department of Educational Psychology
University of Nevada, Las Vegas

Brett D. Campbell

Local Address:
1675 Silver Glen Ave
Las Vegas, NV 89123

Degrees

Bachelor of Arts, Psychology, 1990
Weber State University

Master of Science, Educational Psychology
University of Nevada, Las Vegas

Special Honors and Awards

2002 Graduate Research Training Assistantship. University of Nevada, Las Vegas

2003 Graduate Travel Grant Award


2004 Graduate Travel Grant Award


2002 American Psychology Association Division 15 Graduate Student
Seminar/Mentoring Project participant.

2006 Graduate and Professional Student Association Research Forum
Second place poster in Humanities, University of Nevada, Las Vegas.

2006 Graduate Travel Grant Award

Dissertation Title: The Development of Domain-Specific and Domain-General Monitoring

Dissertation Examination Committee
  Chairperson, Dr. Alice J. Corkill, Ph. D.
  Committee Member, Dr. Gregory Schraw, Ph. D.
  Committee Member, Dr. CarolAnne Kardash, Ph. D.
  Graduate Faculty Representative, Dr. Mark Guadagnoli, Ph. D.