PRESCHOOLERS' PERSISTENT OVERCONFIDENCE
IN THEIR RECALL MEMORY

A dissertation submitted
to Kent State University in partial
fulfillment of the requirement for the
degree of Doctor of Philosophy

by

Amanda Rae Lipko

August, 2008

Dissertation written by
Amanda Rae Lipko
B.A., The College of New Jersey, 2002
M.A., Kent State University, 2005
Ph.D., Kent State University, 2008

Approved by

_____, Co-Chair, Doctoral Dissertation Committee

William E. Merriman

_____, Co-Chair

John Dunlosky

_____, Members, Doctoral Dissertation Committee

Maria Zaragoza

Joel Hughes

Christopher Was

Accepted by

_____, Chair, Department of Psychology

MaryAnn Stephens

_____, Dean, College of Arts and Sciences

John R.D. Stalvey

TABLE OF CONTENTS

LIST OF FIGURES	iv
LIST OF TABLES	v
ACKNOWLEDGMENTS	vi
CHAPTER	Page
1 INTRODUCTION	1
2 EXPERIMENT 1	5
3 EXPERIMENT 2	15
4 EXPERIMENT 3	32
5 GENERAL DISCUSSION	57
REFERENCES	65
APPENDIX	
A PICTURES AND PRESENTATION ORDER USED IN PREDICTION TASKS	72
B WORDS AND PRESENTATION ORDER FOR PSEUDOWORD REPETITION TASK (EXPERIMENT 3)	73
C SENTENCES, EXPECTED WORDS, AND PRESENTATION ORDER USED FOR LISTENING COMPLEX SPAN TASK (EXPERIMENT 3)	74

LIST OF FIGURES

Figure	Page
1 Mean self prediction and recall values across trials of Experiment 1	11
2 Mean self predictions and recall values across trials of Experiment 2	22
3 Mean other predictions and recall values for Experiment 2	25
4 Mean self and other prediction values across trials of Experiment 2	25
5 Mean self predictions and recall values across trials in Experiment 3	44
6 Mean self predictions for think-back and control groups across trials of Experiment 3	44
7 Mean postdictions and recall across trials for the think-back group of Experiment 3	46
8 Mean postdictions and predictions across trials for the think-back group of Experiment 3	48

LIST OF TABLES

Table	Page
1 Intercorrelations Among Prediction and Recall in Experiment 1	13
2 Intercorrelations Among Self-predictions, Recall, and Individual Measures in Experiment 2	26
3 Intercorrelations Among Self-predictions, Recall, and Individual Difference Measures in Experiment 3	50
4 Correlations Between Recall on Trial N and Prediction for Trial N + 1 over Experiments 1, 2, and 3.	50
5 Mean Proportion of Correct Pseudoword Repetitions in Experiment 3	54

ACKNOWLEDGMENTS

There are many people who have helped, supported, and guided me as I worked on this endeavor and I sincerely thank them all. Specifically, I want to thank Bill Merriman and John Dunlosky for their guidance and mentorship on this dissertation and throughout my time in graduate school. I also want to acknowledge the cognitive faculty and my committee members for their assistance and encouragement. I am grateful for the friendship and support of my colleagues and lab-mates. I am indebted to the wonderful staff, children, and parents at the daycare centers at which I collected my data. Without their help, this project could not have been completed.

Most importantly, I am grateful to my parents who unconditionally love and support me. I want to thank my grandparents and my Aunt Martha who is a constant source of encouragement. I also cannot forget to thank my NJ family: Lou, Terry, Megan, Jillian, and Taylor. I especially want to acknowledge my Gigi who is my friend, my role model, and my hero. Finally, I would like to dedicate this dissertation to my late grandfather, Raymond Shipman. He was the first person in my family to earn a college degree and I know he is watching from above with feelings of pride and love.

CHAPTER 1

INTRODUCTION

Young children are overconfident about their abilities. If you spend a day around preschoolers, you are likely to hear them claim that they can accomplish feats that are well above their actual capabilities. Such overconfidence occurs across a variety of physical tasks (e.g., Stipek, Roberts, & Sanborn, 1984; Plumert, 1995, Plumert & Schwebel, 1997 Schneider, 1998) and cognitive tasks (e.g., Bjorklund, Coyle, & Gaultney, 1992; Flavell, Friedrichs, & Hoyt, 1970; Cunningham & Weaver, 1989; Mills & Keil, 2004; Yussen & Levy, 1975; Henry & Norman, 1996; Schneider, 1998; Shin, Bjorklund, & Beck, 2007). Although overconfidence declines with age (Schneider & Presseley, 1997), grade-school students and even adults still show overconfidence in their memory performance (eg. Lipko, Dunlosky, Rawson, Swan, & Cook, 2007; Schneider, Visé, Lockl and Nelson, 2000; Metcalfe, 1998; Plumert, 1995; Plumert & Schwebel, 1997). This overconfidence can lead to poor self-regulation which can be detrimental to individuals' learning goals (Theide, 1999).

The preschool age represents a unique developmental time in which to study overconfidence because preschoolers typically have limited experience in predicting their own abilities. Because they have not yet entered the formal educational system, most have not yet developed strategies for compensating for their overconfidence. Thus,

studying the preschool age group provides researchers with the chance to investigate the mechanisms of overconfidence because across most tasks preschoolers are naturally overconfident.

Early Research Investigating Children's Overconfidence

In a classic study, Flavell, Friedrichs, and Hoyt (1970) used a performance-prediction paradigm in which nursery school, kindergarten, second grade, and fourth grade children were asked to predict their memory span. On each trial, a single picture was added to a strip of paper (up to 10 pictures total), and children were asked if they would be able to remember the pictures once they were covered. Children's predicted memory span was defined as the longest series of pictures they said that they would be able to remember. After completing the predicted memory span task, children's actual memory span was assessed with an aural recall task. An experimenter read a series of familiar object names aloud, and afterwards the children were asked to repeat the names in the correct order. Children's actual memory span was defined as the longest series of words they were able to correctly recall. In all age groups, the children's predicted memory span was higher than their actual memory span. However, the nursery school and kindergarten children gave more overconfident predictions than the 2nd and 4th graders. Fifty-seven percent of the nursery school children and sixty-four percent of the kindergarteners predicted their memory span to be the maximum possible value. Using a similar procedure, Yussen and Levy (1975) also found that preschoolers and third-graders over-predicted their memory abilities but that preschoolers were significantly more overconfident.

Is Children's Overconfidence Adaptive?

In some circumstances, children's overconfidence can be adaptive (Bjorklund & Green, 1992; Shin et al., 2007). Overconfidence can help to maintain and protect children's self-esteem and may serve to encourage children to try new tasks that they may initially poorly perform. Having an optimistic view of one's own abilities can encourage children to try to achieve loftier goals for themselves (Bandura, 1989). If children are motivated to persist at a task, they will subsequently gain practice which itself may lead to increased overall performance (Shin et al., 2007).

However, overconfidence in the face of contradictory task experience can have harmful effects. For example, Plumert (1995) asked 6-year-olds, 8-year-olds, and adults to predict whether they would be able to successfully perform physical tasks that were either within or beyond the participants' actual ability levels. The children who made the most errors in judging their ability to accomplish physical tasks were the same children who had experienced more accidents compared to the children who had made fewer errors in judgments.

Remaining overconfident in the face of feedback can also lead to persistent failure on some tasks, especially if young children believe that the amount of effort they exert on a task is directly predictive of their success (Stipek & Mac Iver, 1989). Believing that effort rather than their previous performance is a good indicator of future performance, children are likely to remain overconfident and may not seek to improve their performance. Additionally, young children are less likely to spontaneously generate and utilize beneficial strategies compared to older children (e.g. Kail, 1990, Kreutzer,

Leonard, Flavell, & Hagen, 1975) and overconfidence may further encourage this tendency. If children are unaware that they need a strategy, they will be even less likely to search for or implement one.

Overview of Current Investigation

The goal of the current investigation was to further examine overconfidence in the preschool age group. More specifically, three experiments assessed whether 4- and 5-year-olds' overconfidence persisted across multiple trials of a memory task or whether children incorporated task experience into their predictions and thus lowered their expectations of their recall performance. In each of the three current experiments, children named ten pictures, and following a brief study time, predicted how many they would be able to recall. After they had actually attempted recall, the children were told how many pictures they had recalled. The procedure was then repeated with a new set of pictures. Three trials were administered in Experiment 1, and five trials in the remaining experiments.

CHAPTER 2

EXPERIMENT 1

Introduction

In early studies of overconfidence (e.g. Flavell et al., 1970; Yussen & Levy, 1975), participants were asked to predict their memory span using pictures and then they were given a recall task with different stimuli. Thus, children were not asked to predict their memory for the actual stimuli they would be attempting to recall. In 1989, Cunningham and Weaver asked 6-year-olds to make predictions about the same words they were asked to later recall. After making a prediction, children listened to a recording of ten words and then attempted to recall those words. Across five trials, the six-year-olds' predictions were overconfident.

However, children's predictions were less overconfident when they were exposed to the stimuli prior to making their predictions. In a concurrent version of the task, children were asked to listen to words and to press the stop button on the tape recorder once they had heard as many words as they thought they would be able to remember. Children's final prediction was less overconfident and almost accurate on the concurrent task whereas it remained overconfident on the prospective task. Schneider (1998) asked 4-and 6-year-olds to engage in the same concurrent task and found that both age groups

overestimated their recall but that the 4-year-olds were significantly more overconfident. Schneider (1998) also asked the children to make predictions about their performance on a memory for location task and found that six-year-olds were more overconfident in their memory than 4-year-olds.

More recently, Shin et al. (2007) asked kindergarteners, first-graders, and third-graders to predict picture recall in a supraspan task. Participants predicted how many of pictures they would be able to recall if they were shown fifteen pictures, then were tested for how well they could actually recall fifteen pictures. As in the previous studies, kindergarteners and first-graders significantly overestimated their recall. They were significantly more overconfident than the third-graders.

Although performance prediction studies like the ones discussed above have demonstrated that young children are overconfident, the degree of children's overconfidence across trials of a task has not been as consistently reported. Investigating change in children's predictions across trials allows for a better evaluation of potential explanations of children's overconfidence. For example, the wishful thinking hypothesis posits that children's predictions are based on how they want to perform rather than on how they expect they will perform (Stipek et al., 1984). If children base their predictions on wishful thinking, their predictions should remain equally overconfident across all trials.

The studies which have provided data on change across trials for children's predictions on a supraspan task yield mixed results regarding whether children change their predictions after repeated experience with the task. Shin et al. (2007) reported that

kindergarteners' predictions did not change across five trials, whereas the kindergarteners tested by Cunningham and Weaver (1989) reduced their predictions after the first trial.

In the latter study, the mean level of children's predictions did not change after the second trial, and exceeded the mean level of actual recall on every trial. Due to a filler task that was administered, there was a longer time delay between stimulus presentation and attempted recall in Shin et al. (2007) and this may explain these discrepant findings. Children may have forgotten the actual stimuli and/or their previous trial recall because of the greater time delay between trials. Thus, at the time of prediction, they may not have based their prediction on their ability to remember the actual stimuli. Rather, their predictions may have been influenced by wishful thinking (Stipek et al., 1984) or their general inflated beliefs about their memory.

One procedure common to the studies discussed above is that children made predictions before they were presented the actual study stimuli. This procedure was also used in the single trial prediction studies of Flavell et al. (1970) and Yussen and Levy (1975). Thus, in these studies, children's predictions were not based on an evaluation of their ability to remember the particular set of items they were asked to study but rather on some more general evaluation. Regarding studies of adults that have used this same type of procedure, Hertzog, Dixon, and Hultsch (1990) have argued that without knowledge of the particular-to-be learned items, judgments cannot be influenced by monitoring of current learning per se, and so are largely based on general beliefs about one's own memory. Adults' predictions made prior to study have been found to be different than

those made immediately after study, with the latter typically being more accurate (Hertzog, Saylor, Fleece, & Dixon, 1994).

Thus, the present experiments examined whether persistent overconfidence occurs when preschoolers have an opportunity to study the list of to-be-recalled items prior to predicting their performance. In particular, a first goal of Experiment 1 was to investigate the accuracy of children's post-study predictions of their picture recall. The second was to evaluate whether this overconfidence could be eliminated with repeated recall experience. To foreshadow, children were overconfident across all of the trials but their predictions declined from the first to last trial. As explained in detail later, Experiments 2 and 3 were designed to directly evaluate reasons why recall experience did not further reduce children's overconfidence in their predictions.

Method

Participants

Twenty-one children ($M = 5-0$, range = 4-0 to 5-11) were recruited to participate. There were twelve girls and nine boys. They were recruited through their preschools which were located in a primarily suburban, middle-class section of Northeastern Ohio.

Materials and Procedure

Children were tested individually in a quiet room in their preschool. Children sat at a table next to the experimenter in front of a large, white foam board that contained ten magnets. Before beginning the task, children were asked a few practice questions to

ensure they understood how to count to ten and could compare quantities of small groups of objects. All children answered these questions correctly.

After the practice questions, the experimenter explained the task to the children by saying, “I am going to put these pictures on the board. I am going to put one picture on each of those magnets. When I am finished, there will be ten pictures on the board. As I put each picture on the board, I want you to tell me the name of what you see in the picture. After all ten pictures are up, I am going to tell you to look at them because shortly after I am going to cover them with a big red piece of paper so you will not be able to see them anymore (the experimenter demonstrated this by placing the paper over the board covering the magnets). But before I cover them, I am going to ask you to tell me how many of the pictures you think you will be able to remember once you cannot see them anymore. After you tell me how many you think you will remember, I am going to cover the pictures and then you are going to try to remember the pictures that are hiding underneath the piece of paper.”

The task began with the experimenter placing ten “4 x 6” colored pictures of familiar objects on the magnetic board (see Appendix A for A list of THE pictures). The pictures were taken from *My First Word Book* (Wilkes, 1999). If an object could not be correctly named, the picture was replaced with another picture. After all ten pictures had been presented the children were given ten seconds to study the pictures. They were then asked, “How many of these pictures do you think you are going to be able to remember once I cover them up?” The pictures were covered with a large piece of red paper and the children were asked to recall as many of the pictures as possible. After sixty seconds,

children were asked to stop trying to recall pictures and the pictures were uncovered.

Most children stopped recalling pictures before this time limit.

After uncovering the pictures, the experimenter told the children the number of pictures they had recalled. The experimenter also removed the same number of pictures from the board as a visual demonstration of this number. However, the actual pictures the children recalled were not necessarily removed. (This methodological decision was based on time constraints. It took longer to remove the exact pictures a child recalled on each trial and it was important for the experiment that the trials occur as close in time as possible). The entire procedure was repeated twice more, each time with ten different pictures. At the end of the third trial, the children were given a small gift of stickers to thank them for their participation.

Results and Discussion

Levels of Prediction and Recall

The mean prediction and recall values for each of the three trials are presented in Figure 1. A 2 (task: prediction vs. recall) x 3 (trial) mixed analysis of variance of children's predictions and recall yielded significant main effects of task, $F(1, 20) = 54.85$, $p < .01$ and trial, $F(2, 40) = 18.11$, $p < .01$. Children's predicted recall ($M = 7.89$, $SD = 2.05$) was substantially higher than their actual recall ($M = 4.27$, $SD = 1.15$) on all three trials. Thus, as in previous studies, young children's overconfidence in their memory was quite pronounced. The majority of children (12 of 21) actually predicted that they would recall all ten pictures on the first trial, yet none of them did so.

Figure 1. Mean self prediction and recall values across trials of Experiment 1.

Both prediction and recall values declined over trials. However, task interacted with the quadratic component of trial, $F(1,20) = 6.90, p < .05$. Recall declined significantly from trial 1 to 2, $t(20) = 4.75, p < .01$, but not from trial 2 to 3, $t < 1$, whereas prediction did not change from trial 1 to 2, $t < 1$, but declined significantly from trial 2 to 3, $t(20) = 2.62, p < .05$. Thus, the decline in prediction lagged behind that in recall. The pattern of decline in recall matches the typical pattern observed in studies of proactive interference in children's recall (Kail, 2002).

These results are more similar to those of Cunningham and Weaver (1989), in which a decline over trials in prediction and recall was observed, than those of Shin et al. (2007), in which only recall declined over trials. The match with Cunningham and Weaver's results were not exact, however. In their study, predicted recall declined from trial 1 to 2 whereas it declined from trial 2 to 3 in the present study. The reasons for this discrepancy are not clear. Children may need more experience with a picture recall task compared to an aural recall task, which is what Cunningham and Weaver used, before lowering their predictions. It is also possible that children are more likely to delay reducing their predictions when the to-be-remembered items are presented before predictions are made, as in the current study, rather than after predictions are made, as in Cunningham and Weaver's study. It may be easier for children to imagine how they might recall more on trial 2 than on trial 1 when they are shown the new set of pictures to be used on trial 2 compared to when they are not shown the stimuli.

Relationships between Prediction, Recall, and Age

Intercorrelations are presented in Table 1. Stable individual differences were evident in predicted recall (cross-trial r 's between predictions ranged from .48 to .63) and to a lesser degree, in actual recall (cross-trial r 's ranged from .34 to .48). Children's predicted recall for a trial did not correlate with the number they had just recalled on the previous trial suggesting that children were not influenced by their previous recall when making their subsequent predictions.

Table 1. *Intercorrelations Among Prediction and Recall in Experiment 1.*

	P1	P2	P3	R1	R 2	R3
Age	.31	-.04	-.03	.14	.32	.09
Prediction 1		.57**	.48*	.12	.23	-.09
Prediction 2			.63**	.08	.14	.12
Prediction 3				.27	.04	-.24
Recall 1					.34	.48*
Recall 2						.36

Note: * $p < .05$; ** $p < .01$

One finding was consistent with the possibility that change in prediction was related to change in recall, however. Those children who lowered their predicted recall for the final trial by the greatest amount tended to show the least decline in recall from previous trials to the final trial, $r(19) = -.41$, $p = .06$; with age partialled out, $r(18) = -.46$, $p < .05$. Those individuals who realized after two recall trials that they were not going to

recall as many pictures as they had originally thought, and so lowered their prediction, may have been prompted by this same realization to examine the next set of pictures more closely and/or extensively resulting in improved recall, or at least greater resistance to proactive interference.

Although less relevant to the present concerns, other correlations in Table 1 are informative to the larger literature on children's metamemory performance (Schneider & Lockl, in press). In particular, correlations between predictions and recall on a given trial provide a between-participant measure of judgment accuracy. For the preschoolers in Study 1, these correlations were consistently low and not statistically different from zero, indicating poor monitoring accuracy within a trial. Also, note that age was not significantly correlated with predictions, suggesting that developmental trends within this small range of ages are not robust on these measures of metamemory and memory.

CHAPTER 3

EXPERIMENT 2

Introduction

In Experiment 1, preschoolers were overconfident across all three recall trials. Although predictions declined from trial 2 to 3, the mean prediction on the final trial was still substantially overconfident. One hypothesis that has been proposed for children's overconfidence on prediction tasks is wishful thinking (Stipek et al., 1984). According to this hypothesis, children base their predictions on how they desire to perform rather than on how they think they will actually perform. The main goal of the second experiment was to evaluate this hypothesis as an explanation for children's persistent overconfidence in their picture recall. A secondary goal of Experiment 2 was to investigate possible explanations for variations in children's prediction patterns across trials.

The wishful thinking hypothesis has been evaluated in both psychomotor and cognitive tasks. Stipek et al. (1984) asked 4-year-olds to predict how well they themselves or another child could use a pulley to move a cart up a tower without a marble on the cart falling off. They tended to over-predict their own performance by a greater margin than they did for that of another child. This result is consistent with the wishful

thinking hypothesis under the assumption that children's desired level of performance for themselves exceeds that which they desire for other children.

Schneider (1998, Experiment 1) replicated Stipek et al.'s (1984) findings with a task in which 4- and 6-year-olds predicted how many balls they would shoot into a basket. Although 6-year-olds tended to make more accurate self- and other-predictions than 4-year-olds, the extent to which self-predictions exceeded other-predictions was comparable in the two groups. In further support of the wishful thinking hypothesis, the number of shots that one group of children predicted they would make on each trial was comparable to the number that another group of children indicated they *wished to make* on each trial. This last finding was also obtained in a multi-trial task in which children indicated on each trial how far they either expected or wished to jump, and then attempted to jump as far as they could.

In a second experiment, Schneider (1998) addressed these same issues regarding 4- and 6-year-olds' metamemory. On each test trial, children had to recall where each of ten pictures had been hidden. The pictures changed locations from trial to trial. The children also observed another child perform this task and predicted how well this child would do on each trial. In contrast to the results for the physical skill tasks, children over-predicted the other child's performance by the same margin as they over-predicted their own performance.

These results are open to several interpretations. They may indicate that the wishful thinking hypothesis for memory overestimation is invalid, at least for location memory. Secondly, as Schneider speculated, the fact that the other child was getting

fewer picture locations correct than predicted may not have been very salient. The observing children's attention may have been focused on the task itself, such as on trying to remember the correct locations of the pictures that the other child was attempting to find. At the end of each trial, the observing child was told how many picture locations the other child had identified correctly, however. Thirdly, children may not necessarily desire better performance for themselves in this task than they desire for other children.

One other finding from the experiment did support the wishful thinking hypothesis. The mean number of pictures that children indicated that they expected themselves to remember on each trial was no different from the mean number that other children indicated that they *wished to* remember on each trial. This last finding was also obtained in a memory span task.

In Experiment 2, self- and other-predictions were compared for a multi-trial picture recall task that was similar to the one used in Experiment 1. If the wishful thinking hypothesis is valid and it is also true that children desire better picture recall performance for themselves than for another child, then participants should over-predict their own recall performance by a greater margin than they over-predict the recall performance of another child.

Two trials were added, for a total of five trials. Because children's predictions only began to decline after the second trial in Experiment 1, and were still well above actual performance by the third trial, the goal was to provide more recall experiences. It was hypothesized that participants' predictions would continue to decline after the third

trial if they tended to infer from the experience of repeated under-performance that they needed to continue to lower their expectations for performance.

Although children's predictions remained overconfident on the final trial of Experiment 1, individual patterns of prediction and recall across trials varied. Those who lowered their prediction for the final trial by the greatest margin tended to show the least decline in recall over trials. It was hypothesized that children who lowered their predictions the most after trial two were also the most likely to realize that they needed to increase their efforts to study the pictures for recall. Because they then increased these efforts on the later trials, they experienced less decline in recall over trials than the other children.

To shed further light on these individual differences, Experiment 2 included assessments of children's performance on a vocabulary test and a test of their understanding of number (Okamoto & Case, 1996). Children's vocabulary often serves as a good measure of general intelligence. It is possible that the different prediction patterns exhibited by the children are a result of differences in overall intelligence (Bjorklund, Coyle, & Gaultney, 1992). These patterns may also be explained by differences in children's understanding of the relative cardinal value of numbers (eg. that "five" denotes a larger set of things than "seven"). Children who understand number may make better calibrated yet still overconfident predictions on the memory task. More importantly, children who do not understand the relative cardinal value of numbers may not be able to successfully incorporate previous trial performance into their predictions.

Method

Participants. Thirty-two children ($M = 4-8$, range = 4-0 to 5-4) were recruited to participate. None of the children had participated in the first study. There were twenty girls and twelve boys. They were recruited through their preschools which were located in a primarily suburban, middle-class section of Northeastern Ohio.

Design, Materials, & Procedure

A 2 (prediction: self versus other) X 2 (order: self prediction first versus self prediction second) X 2 (gender of observed child: female vs. male) mixed design was used. The first factor was manipulated within participants. The other factors were counterbalanced and manipulated between participants.

Children were tested individually in a quiet room in their preschool. Before beginning the task, they were asked a few practice questions to ensure that they understood how to count to ten and could compare quantities of small groups of objects. After the practice questions, the experimenter explained the task to the children. Children participated in two conditions of the picture memory task, self prediction and other prediction. The materials were identical in both conditions. After completing both conditions, children were given a small reward of stickers.

Self Prediction Condition

The materials and procedure were identical to those used in the first study with one exception. Two additional trials each with ten new pictures were added to the task for a total of five trials (see Appendix A for list of pictures).

Other Prediction Condition

Children were told that they were going to watch a DVD of a similarly aged boy or girl playing a memory game. On the DVD, either a male or a female preschooler (the model) performed five trials of the self prediction task with the experimenter. Each trial began with the model looking at ten pictures on a magnetic board. The experimenter paused the video, then pointed to and named each of the pictures for the participant, and asked, “How many pictures do you think Adam (Morgan) is going to be able to remember once the pictures are covered up?” After giving their prediction, the participant watched the model attempt to recall the pictures. The model had been instructed and prompted to recall pre-determined pictures. Over the five trials, the model recalled 4, 4, 4, 3, and 3 pictures, respectively. This pattern of recall was decided upon based on the average level of recall across trials in Experiment 1. After observing the model’s recall, the experimenter told the participant how many pictures Adam or Morgan had been able to recall and then the next trial began.

Understanding of Cardinality (Okamoto & Case, 1996)

Children judged which of two numbers was bigger or smaller. The first question was, “Which of these two numbers is bigger, 5 or 4?” This question was repeated for the choices, 7 or 9. They children were then asked, “Which of these two numbers is smaller, 8 or 6?” This question was repeated for the choices, 5 or 7.

Vocabulary

A modified version of the vocabulary subtest of the Wechsler Preschool and Primary Scale of Intelligence – 3rd edition (Wechsler, 2002) was administered. Children were asked to explain the meanings of twenty words. Only the verbal items of the subtest were used and the test was stopped once a child incorrectly defined five words in a row. Each word was scored either as a 0, 1, or 2 following the standard scoring criteria. The maximum possible score was 38.

Results and Discussion

Analyses of mean self-prediction, other-prediction, and recall (by self) will be presented first. Self- and other-predictions will then be compared to address whether these are consistent with the wishful thinking hypothesis under the assumption that children desire better memory performance for themselves than for other children. Next, intercorrelations among these measures will be examined. Finally, relationships between the individual difference measures and the prediction and recall measures will be explored.

Levels of Prediction and Recall

The mean self-prediction and recall values for each trial are presented in Figure 2. A 2 (task) x 2 (order) x 5 (trial) mixed analysis of variance of children's self-predictions and recall scores yielded significant main effects of task, $F(1,30) = 76.76, p < .01$ and trial, $F(1, 30) = 13.99, p < .01$. As in Experiment 1, the level of recall that children

Figure 2. Mean self predictions and recall values across trials of Experiment 2.

expected to achieve on every trial (overall $M = 6.18$, $SD = 2.11$) was substantially higher than what they actually achieved on that trial (overall $M = 2.81$, $SD = .98$).

Both self-predicted and actual recall declined over trials, replicating another finding from Experiment 1. The mean prediction of the first trial was significantly different from the mean prediction on the fifth trial, $t(31) = 2.53$, $p < .05$. The mean prediction on the fourth trial was also significantly different from that on the fifth trial, $t(31) = 2.06$, $p < .05$. No other differences in prediction were statistically significant, $ts < 1$. The mean recall on the first trial was significantly different from the mean recall on the second trial, $t(31) = 4.51$, $p < .01$. No other differences in recall were statistically significant, $ts < 1$.

As in Experiment 1, the decline in prediction lagged behind that in recall. Task interacted with the quadratic trend component of trial, $F(1, 30) = 6.64$, $p < .05$. The extent to which the decline in prediction lagged behind that in recall was greater in the current experiment (3 trials later) than in Experiment 1 (1 trial later). The source of this discrepancy may be the experience that was unique to participants in the current experiment, namely, of observing another child perform the task and making predictions about his or her performance. If only the data of those who made self-predictions before other-predictions are considered, self-predictions did drop from trial 2 to 3 by a margin ($M = 1.00$) that was not significant, but was very similar that obtained in Experiment 1 ($M = 1.24$), $t(35) < 1$.

Comparison of Self- and Other-Predictions

Mean other-predictions for each trial are presented in Figure 3. These predictions were significantly greater than the observed child's modeled recall on every trial (i.e., greater than 4, 4, 4, 3, and 3, respectively). The least overestimation was observed on the first trial, but even this discrepancy (M difference = 2.62 items) was significant, $t(31) = 4.54, p < .001$.

In a 2 (type of prediction) x 2 (order) x 5 (trial) mixed analysis of variance of self- and other-predictions, the only significant effect was the linear trend component of trial, $F(1, 30) = 4.20, p < .05$. There was no evidence that self- and other-predictions differed (see Figure 4). Both exceeded actual recall by similar magnitudes, and both declined over trials in a similar manner. These results replicate those of Schneider (1998, Experiment 2) with a picture recall task rather than location recall task. To accommodate these findings, it may be necessary to reject either the wishful thinking hypothesis for children's memory overestimation or the assumption that children desire better memory performance for themselves than for other children.

Relationships between Age, Self-Prediction, and Recall

Intercorrelations are presented in Table 2. As in Experiment 1, stable individual differences were evident in self-predictions (median cross-trial r was .46) and to a lesser degree, in actual recall (median cross-trial r was .34). Also, as in Experiment 1, self-predictions were not significantly related to age on any trial. However, in contrast to the results of Experiment 1, recall on certain trials was significantly correlated with age. Given the large number of correlations that were computed, some may be significant by

Figure 3. Mean other predictions and recall values for Experiment 2.

Figure 4. Mean self and other prediction values across trials of Experiment 2.

Table 2. *Intercorrelations Among Self-predictions, Recall, and Individual Difference Measures in Experiment 2.*

	2	3	4	5	6	7	8	9	10	11	12	13
1. Age	.24	.03	.25	-.07	.11	.34	.46*	.26	.11	.40*	.40*	.43*
2. Prediction 1		.42*	.54*	.19	.41*	.44*	-.05	.08	.15	.06	.17	.10
3. Prediction 2			.35*	.39*	.43*	.26	-.10	-.13	.07	-.20	.42*	.04
4. Prediction 3				.53**	.63**	.29	.00	.20	.22	-.13	.04	.27
5. Prediction 4					.77*	.11	-.22	.19	.40*	-.03	-.13	.13
6. Prediction 5						.19	-.04	.29	.41*	-.03	.12	.28
7. Recall 1							.25	.50**	.21	.27	.25	.38*
8. Recall 2								.50**	.31	.30	.21	.48**
9. Recall 3									.35	.44*	.19	.40*
10. Recall 4										.25	-.17	.06
11. Recall 5											.23	.22
12. Vocabulary												.41*
13. Cardinality												

Note: * $p < .05$; ** $p < .01$

chance. However, all five correlations between recall on a particular trial and age were positive, with median $r = .34$, and total recall was also significantly correlated with age, $r(30) = .45, p < .01$. Several other studies have found older preschoolers to recall more pictures than younger ones (see Perlmutter & Lange, 1978; Schneider & Pressley, 1997). The discrepancy between the results of Experiments 1 and 2 regarding the relation between age and recall may just be due to sampling error. Although the correlation between age and total recall in Experiment 1 was not significant, it was positive, $r(19) = .23, p > .10$.

Despite the evidence that recall increased with age and that the level of recall that children predicted for themselves did not, it cannot be concluded that over-prediction errors were smaller in older preschoolers than in younger ones. Age was not significantly correlated with the difference between self-predicted recall and actual recall on any particular trial or in the aggregate. This paradox may be due to a problem of measurement. Because a difference score is affected by the noise in each of the scores that make it up, it tends to be less reliable than its component scores (Cronbach, 1997). The less reliable a score, the less statistical power any particular study has for detecting significant relations between it and other variables.

In Experiment 1, the number of pictures that children predicted they would recall on a particular trial was not correlated with the number they actually recalled on that trial or the previous trial. However, three correlations of this type were significant in the current experiment. The number of pictures that children recalled on trial 1 was positively correlated with the number that they predicted they would recall on that trial.

Also, recall on trial 4 was positively correlated with self-predicted recall for both that trial and the next one. Again, because of the chance of getting a few Type I errors when computing a large number of correlations, caution is necessary in interpreting these results.

The correlation coefficient for the relation between self-prediction and recall on the first trial of Experiment 1 (.12) was actually not significantly different from the one obtained in the current experiment (.44), $Z = 1.41$, $p > .10$. So the discrepancy between these results for this particular relation may just be due to sampling error. For the combined results of the experiments (using the Stouffer method, Rosenthal, 1991), the relation between self-prediction and recall on the first trial was significant, $Z = 2.14$, $p < .05$, mean Fisher $Z_r = .30$, which corresponds to $r = .29$. So support was obtained for the claim that those who predict higher levels of recall on the first test actually tend to recall more than other children on this test.

The positive relation between self-prediction and recall was not maintained after the first trial. Same-trial correlations between these two measures were not significant for the second or third trial in either the current experiment or Experiment 1. Children who predicted higher levels of recall on the first test experienced sharper declines in recall after the first test. Self-predictions for the first trial were correlated with the difference between recall on that trial and average recall on the next two trials, $r(30) = .45$, $p < .005$. In Experiment 1, this correlation was not significant, $r(19) = .08$. For the test of this particular difference, $Z = 1.80$, $p = .07$, which indicates that the chance of it only being a reflection of sampling error is rather low (i.e., .07).

Magnitude of decline in recall may be positively associated with the level of first trial prediction only when the latter prediction is positively associated with first trial recall. The combined results of the two experiments are consistent with this conjecture. For these data, self-prediction and recall on the first trial were related (as already reported), and self-prediction for the first trial was correlated with the decline in recall from the first trial to the next two trials, $Z = 2.27$, $p < .05$, mean Fisher $Z_r = .28$, which corresponds to $r = .27$. This finding presents a challenge for the proposal that young children with higher expectations than their peers will tend to show smaller declines in memory performance over trials (Shin et al., 2007; Bjorklund et al., 1992).

Since, Experiment 1 only included three trials, the current finding of significant positive relations between recall on the fourth trial and predictions on the fourth and fifth trial were unique to the current experiment. It is not clear why children who predicted higher recall than their peers on the fourth trial actually recalled more than their peers thereafter, but no relation between prediction and recall was evident on the two preceding trials. Perhaps those who made very high predictions on the fourth trial, despite having experienced three trials in which they recalled much less than they predicted, tended to persist more in their study and retrieval efforts on trials four and five. The other children may have become discouraged after three trials of less-than-expected-recall, and so began to predict less recall, reduced their efforts during study and retrieval, and consequently, began to recall fewer items.

This interpretation is generally consistent with the proposals of Bjorklund and colleagues regarding the motivating effects of high expectations in young children.

However, it does not match the way that Shin et al. (2007) interpreted these proposals for their data. They took the proposals to imply that those who showed the highest expectations on the first two trials would show the least decline in recall over trials. This prediction was not upheld in the current study, as the level of initial self-predictions (i.e., each child's average prediction on trials 1 and 2) was not related to how much recall declined from trials 1 and 2 to trials 4 and 5; in fact, the trend was in the opposite direction, $r(30) = .19, p > .10$.

Vocabulary

Children scored an average of 11.9 points ($SD = 3.74$) of the possible thirty-eight points. Vocabulary performance was only significantly related to the mean self prediction performance on Trial 2, $r = .43, p < .05$ (see Table 2). It was not significantly related to mean self prediction performance on any other trial or to mean recall performance on any trial. Vocabulary size does not appear to explain the individual differences in children's self prediction patterns.

Understanding of Cardinality

Sixteen of the thirty-two children performed at chance or below on the cardinality test ($M = .39, SD = .16$) whereas sixteen answered at least three of the four questions on this test correctly ($M = .83, SD = .12$), $t(30) = -8.85, p < .01$. Understanding of cardinality was not significantly related to self prediction performance (see Table 2). However, it was significantly related to mean recall performance on Trials 1, 2, and 3, as well as to total recall ($r(30) = .45, p < .05$). Contrary to expectation, children's

understanding of cardinality does not appear to explain the individual differences in children's self prediction patterns.

CHAPTER 4

EXPERIMENT 3

Introduction

In the first two experiments, preschoolers overconfidently predicted their picture recall. This overconfidence persisted across multiple trials despite children's repeated experience of recalling fewer pictures than predicted. According to the wishful thinking hypothesis (Stipek et al., 1984; Schneider, 1998), children overestimate because they base their predictions on how they desire to perform rather than on how they believe they will perform. This hypothesis was not supported by the results of Experiment 2. Children's predictions of how another child would perform on the picture recall task were no different from their predictions of how they themselves would perform. This finding replicates and extends Schneider (1998) using a different memory task. Wishful thinking appears to be a viable explanation for children's overconfidence on psychomotor tasks (Stipek et al., 1984; Schneider, 1998) but does not appear to explain children's overconfidence in memory tasks. However, there is still a possibility that children are equally influenced by wishful thinking when predicting their own performance as well as when they are predicting another child's performance on a memory task.

Alternative Explanations for Children's Overconfidence

An alternative explanation for children's overconfidence on a memory task is the memory monitoring deficiency hypothesis (Schneider, 1998). According to this hypothesis, young children do not accurately monitor their performance on a task. As a result, they continue to give predictions that are well above their demonstrated performance level. Schneider evaluated this hypothesis by asking 4- and 6-year-old children to report their final trial's performance in a location memory task. These postdictions were quite accurate, suggesting that children monitored their memory performance accurately.

If children are not engaged in wishful thinking and they are accurately monitoring their performance, why does their overconfidence persist across trials? In a multi-trial task, monitoring previous performance is only one part of making an accurate prediction. It is also essential to use this information to temper one's overly optimistic expectations. In Experiments 1 and 2 of the current investigation, children did not appear to incorporate outcomes from previous trials into their predictions as indicated by the null correlations between recall performance on Trial N-1 and predictions on Trial N (see Tables 1 and 2). One exception was on Trial 5 in Experiment 2, which may be evidence that task performance that is less than expected may help to lower children's expectations for future performance. Even if this adjustment occurred in Experiment 2, it was far from adequate. Children's predictions on the final trial were substantially overconfident.

The use of one's own past performance in predictions of one's future performance is not always an easy feat. Even adults do not fully use information from task experience

when predicting their memory performance (Dunlosky & Hertzog, 2000). Additionally, being aware of one's previous experience does not guarantee that one will be influenced by such information (Pressley & Ghatala, 1989). For example, on a psychomotor tower task, Stipek et al. (1984) reminded children of their previous performance immediately before asking them to predict their performance on the next trial. The reminder consisted of directing children to look at the number of colored tacks displayed on the tower, which represented how high they had been able to maneuver the cart on the previous trial. The experimenter also moved the cart to the final level the child had reached as another demonstration of their previous performance. Such reminders did not influence the preschoolers' predictions.

Similar difficulties surround the use of other individual's experience with a task. Yussen and Levy (1975) found that preschoolers' predictions of their own memory span performance were not influenced by information that they were provided about the typical performance of similarly-aged children on the same task. Children's predictions in Experiment 2 of the current study were also not influenced by a model child's recall performance. Even after observing another child complete five trials of the picture memory task, children remained overconfident in their predictions.

Even if children can monitor accurately and can understand the benefits of using task experience in the formulation of a new performance prediction, they still may not spontaneously think about such information and/or incorporate it into their predictions of future performance (e.g. Kail, 1990). Pressley, Ross, Levin, and Ghatala (1984) asked ten- and thirteen-year-olds to learn vocabulary words using one of two strategies. In all

conditions, children were asked to select one strategy to use during the learning process. Children studied the words and were immediately given a recall test. Some children were given practice with the strategies before making their selection. Children who were given practice and asked to think about the effectiveness of each strategy in their learning of the vocabulary words were more likely to select the more useful of the two strategies compared to children who were given practice but who were not asked to explicitly think back about their experience with the strategies. Importantly, it was determined that regardless of condition, children were aware of the efficacy of both strategies. Although children had knowledge about the usefulness of the two strategies, they did not use this knowledge in their selection.

Overview of Current Experiment

In Experiment 3, one group of children was asked to think back to their previous trial performance immediately before predicting their recall on the next trial. More specifically, they were asked to make a postdiction for their previous trial's performance at the start of each trial. Performance in this think-back group was compared to the performance of a control group who did not make postdictions prior to their predictions.

This design allowed for an evaluation of the two essential components of adaptive performance monitoring: accuracy and influence over prediction. Based on Schneider's (1998) previous findings, it was hypothesized the children's postdictions would be accurate or at least close to accurate. Although some children may not successfully monitor their performance, their postdictions, regardless of accuracy, still allow for the evaluation of their use of such monitoring information in the formulation of predictions.

It is unclear whether or not children's predictions will be influenced by their postdictions. Previous researchers have not examined the relationship between children's postdictions and their subsequent predictions on multiple trials of a memory task. By making postdictions, information about previous performance may become salient enough to influence children's predictions. When Stipek et al. (1984) made outcome information on the psychomotor tower task more salient by dropping a single marble into a glass dish for every interval on the tower that the cart had been maneuvered on the previous trial, children responded by making lower predictions. Thus, increasing the salience of performance outcomes may promote children's use of such information in their predictions.

Individual Differences and Children's Prediction Patterns

Although predictions remained overconfident on the final trials of both Experiment 1 and 2, the pattern of children's predictions across trials varied. Experiment 2 attempted to explain these patterns by assessing children's performance on two measures of individual differences: vocabulary and understanding of number (Okamoto & Case, 1996). Unfortunately, neither measure provided a consistent explanation for these patterns. The current experiment examined another possible explanation for these patterns, working memory capacity, which is a system that allows for the simultaneous storing and manipulating of cognitive information (Baddeley & Hitch, 1974). Researchers have found links between working memory and a variety of cognitive and learning abilities (Alloway, Gathercole, Willis, & Adams, 2004).

Children's persistent overconfidence in their memory is likely the result of more than one causal factor, but working memory may help to partially explain it. Children's limited working memory capacity may prevent them from successfully benefiting from monitoring information and so cause them to remain overconfident across trials. For example, children who recall four pictures on a trial but predict that they will recall 8 pictures on the subsequent trial may not be able to simultaneously think about their previous recall while predicting their next recall. They do not have their previous recall value active or available at the time of prediction and thus their prediction would not be expected to be influenced by their previous performance. Because the structure and capacity of children's working memory develops and changes with age (e.g. Case & Okamoto, 1996; Gathercole, Pickering, Ambridge, & Wearing, 2004; Towse & Hitch, 2007) older children may be better able than younger ones at successfully holding task experience in working memory while formulating a prediction. Although, older children did not differ from younger children in magnitude of overconfidence in Experiments 1 or 2, working memory differences may only matter when performance outcome information is made salient to the child (as in the think-back group of Experiment 3). If performance outcome is salient enough for children, those with better working memory abilities may be better able to retain and use the outcome information in the formulation of their next prediction.

The picture memory task used in the first two experiments required children to name pictures, briefly study them, make a prediction, and then attempt recall. Following recall, children were told how many pictures they remembered. The next trial began with

the naming of ten new pictures. Thus, children needed to maintain performance monitoring information while simultaneously naming and storing the pictures of the next trial before making their next prediction. This procedure is likely to have placed a strain on working memory. It was hypothesized that children who performed well on tests of working memory would demonstrate the most advanced prediction patterns across trials. More specifically, children with high working memory scores were expected to show less overconfidence overall and to make more calibrated predictions across trials compared to those with low working memory scores.

Two working memory tasks were administered, pseudoword repetition (Gathercole, 1995; Merriman et al., in press) and a complex listening span task (Daneman & Carpenter, 1980; Siegel & Ryan, 1989; Leather & Henry, 1994). The pseudoword repetition task assessed the efficiency of children's storage of novel word forms in phonological short-term memory. The phonological short-term memory store is part of the phonological loop, one of the main components of Baddeley's working memory model (Baddeley, 2000). The phonological loop briefly stores verbal information, which can include both words and digits. In the pseudoword repetition task, children were asked to repeat back high wordlike nonwords (as previously judged by adults in Merriman et al., in press) that were aurally presented. Children with greater storage capacity should be able to correctly repeat back more and longer nonwords than children with lesser storage capacity.

The complex listening span task was originally created by Daneman and Carpenter (1980) in order to measure both memory storage capacity and memory

processing capacity. A traditional memory span task only measures storage capacity (e.g. how many words can you remember from a list). In Daneman and Carpenter's original complex span task, adults listened to or read sets of sentences of varying length and then attempted to recall the final words of each sentence in the set. In an adapted version of Daneman and Carpenter's task, children listened to a set of sentences and were asked to fill in the missing final word. After listening to and completing all of the sentences in a set, children were asked to recall in order all of the final words of that particular set (Siegel & Ryan, 1989; Leather & Henry, 1994). Children with greater storage capacity and greater processing capacity should be able to correctly fill in and then later recall the final words of longer sets of sentences.

Additionally, the understanding of cardinality task was also administered to determine whether children's understanding of cardinality might be related to their ability to recollect the number of pictures they recalled on a trial (e.g their postdiction accuracy) and because only half of the sample's self predictions could be used in the analysis in Experiment 2.

Method

Participants

Sixty-four children ($M = 4-11$, range = 4-0 to 5-11) were recruited to participate. None of the children had participated in either of the previous experiments. There were 29 girls and 35 boys. They were recruited through their preschools which were located in a primarily suburban, middle-class section of Northeastern Ohio.

Design, Materials, and Procedure

Children were tested individually in a quiet room in their preschool. Before beginning the tasks, they were asked a few practice questions to ensure that they understood how to count to ten and could compare quantities of small groups of objects. After the practice questions, the experimenter explained the tasks to the children. The self prediction memory task was completed first followed by the pseudoword repetition task, the cardinality task, and the listening span task. At the end of the session, children were given a small gift for their participation.

Self Prediction Memory Task

The materials were identical to those used in the Experiment 2 (see Appendix A for list of pictures). The procedure was the same as the one used for the self prediction task in Experiment 2 but with two important differences. First, children were randomly assigned to one of two groups: Think-back and Control. Second, they were asked a question immediately before predicting their recall on trials 2, 3, 4, and 5. This question differed according to group. In the think-back group, the children were asked, “How many pictures did you remember last time?” In the control group, they were asked, “Do you remember what we did last time?” In both groups, immediately after the children answered the pre-prediction question, the experimenter said, “Now, let’s talk about these pictures,” pointed to the pictures on the board and asked, “How many pictures do you think you will be able to remember once I cover them up?”

Pseudoword Repetition

Originally created by Gathercole (1995) and adapted by Merriman et al. (in press), children were asked to repeat back real words and nonsense words that were two, three, and four syllables in length (See Appendix B). The nonsense words had been previously judged by adults to sound like real words (Merriman et al., in press). A practice word (chair) was presented and if a child did not correctly repeat it or said it incorrectly, an additional practice trial was presented. Twenty-one test words were presented every three seconds. The child had to repeat the word within that interval. A repetition was scored as incorrect if any of its phonemes deviated from the model. Only the nonsense words were scored; the real words served as filler items.

Listening Complex Span Task

Based on a task originally created by Daneman and Carpenter (1980) and adapted by Leather and Henry (1994), children listened to simple sentences for which they had to fill in the missing final word. All of the target words were pre-determined (See Appendix C). The sentences were presented in sets that increased in length from two to five. Each set was composed of three lists of sentences. After listening to and filling in the missing final word for one sentence and then for a second sentence, children were asked to recall the two sentence-final words. This procedure was repeated with two other lists of two sentences, then with three lists of three sentences, three lists of four sentences, and so on. The task continued until a child failed every list of a particular length, where failure was defined as not recalling every sentence-final word in the list. For example, children who successfully remembered at least one complete list in the two-sentence set would move

on to try the three-sentence set. If they could not remember at least one complete list in the three-sentence set, the task would be terminated and they would not move on the four-sentence set.

Scores were calculated based on the highest completed set (a completed set was defined as one in which a child correctly recalled all of the missing words from at least one list in the set) plus the proportion of the remembered items from the set immediately after the highest completed set. For example, a child who successfully completed at least one of the lists in the two-word sentence set and remembered 3 out of the 9 words in the three-word sentence set would be given the score of 2.33 (2 points for the completion of the two-word set and .33 for the 3/9 words remembered in the three word list).

Understanding of Cardinality (Okamoto & Case, 1996)

Materials and procedure were identical to those used in Experiment 2.

Results and Discussion

Analyses of mean prediction and recall will be presented first. Predictions in the think-back group will then be compared to predictions in the control group to address whether having children make postdictions prior to making predictions influenced their levels of overconfidence. Next, the accuracy of children's postdictions and their influence on later predictions will be examined. Finally, relationships between the individual difference measures and the prediction and recall measures will be explored.

Levels of Prediction and Recall

A 2 (task: prediction vs. recall) x 5 (trial) x 2 (group: think-back vs. control) mixed analysis of variance of children's predictions and recall yielded significant main effects of task, $F(1, 61) = 176.01, p < .01$ and trial, $F(4, 61) = 27.47, p < .01$. As in the previous two experiments, children's predicted recall on a trial (overall $M = 7.61, SD = 2.29$) was substantially higher than what they actually recalled on that trial (overall $M = 3.56, SD = 1.00$).

Self-predicted and actual recall declined over trials, replicating another finding from Experiments 1 and 2 (see Figure 5). Mean prediction on the first trial was significantly greater than the mean prediction on each of the other four trials, $t_s > 2.64, p < .05$. Mean prediction on the second trial was also significantly greater than that on the third trial, $t(62) = 2.27, p < .05$. No other differences in prediction were statistically significant, $t_s < 1$.

Mean recall on the first trial was significantly greater than the mean recall on each of the other four trials, $t_s > 6.72, p < .01$. The mean recall on the second trial was also significantly greater than the mean recall on the fourth trial, $t(63) = 2.62, p < .01$. No other differences in recall were statistically significant, $t_s < 1$.

Contrary to expectation, the effect of group was not significant, $F(1, 61) < 1$. Similar to the results of Experiments 1 and 2, predictions in both the control and think back groups were overconfident across all five trials (see Figure 6). Thus, despite being asked to make a postdiction immediately before making a prediction for the next trial, children in the think-back group showed persistent overconfidence.

Figure 5. Mean self predictions and recall values across trials in Experiment 3.

Figure 6. Mean self predictions for think-back and control groups across trials of Experiment 3.

Accuracy of Postdictions

A postdiction accuracy score was calculated by dividing the number of correct postdictions by the total number of postdictions for each child in the think-back group. Scores ranged from 0 to 1 with 0 representing no postdiction accuracy and 1 representing perfect postdiction accuracy. The mean level of postdiction accuracy was .74 ($SD = .33$). Overall mean level of recall and postdiction across trials is presented in Figure 7.

Accuracy scores were bimodally distributed. Twenty-three children made accurate postdictions on at least three of the four trials ($M = 3.73$), whereas the other ten did so on two or fewer ($M = 1.26$). This result replicates Schneider (1998) which reported that preschoolers could accurately monitor their performance on a memory task.

It is important to note that Schneider did not report children's overall level of postdiction accuracy so it is unclear how similar the postdiction accuracy found in the current experiment is to that found in his study. Schneider also did not compare children who provided accurate postdictions to those who did not. Additionally, his participants were only asked to make postdictions on the final trial of a task whereas in the current experiment children made postdictions on four of the five trials.

When children in the current experiment did make incorrect postdictions, their recollected recall tended to exceed their actual recall. Among the ten children who made at least two postdiction errors, this inflation tendency was substantial ($M \text{ Error} = + 4.06$ items). Among the six who made only one error, it was not ($M \text{ Error} = + .50$). Postdiction accuracy did not vary over trials. For Trials 1 through 4, mean error (e.g. postdiction – recall) in the think-back group was 1.00, .94, .69, and .79, respectively.

Figure 7. Mean postdictions and recall across trials for the think-back group of Experiment 3.

There was a trend toward a decline in magnitude of error over trials in the ten least accurate children (M Error = 3.36, 3.30, 2.45, 2.40, respectively), but it was not significant, $F(1, 9) = < 1$ for the linear trend component. Because on rare occasion a child's postdiction was lower than actual recall, these last analyses were recomputed with mean error set equal to the absolute value of prediction minus recall. Results were essentially the same.

The prediction and recall performance of participants who made accurate postdictions on every trial ($N = 17$) was compared to that of those who made at least one error ($N = 16$) in a 2 (posdiction accuracy: accurate vs. inaccurate) x 2 (task: prediction vs. recall) x 5 (trial) ANOVA of children's prediction and recall. Postdiction accuracy was not a significant main effect, $F(1,30) < 1$, and was not part of any significant interactions. The children who made accurate postdictions on every trial showed substantial overprediction (M Error = 3.89), which was not significantly less than that of the other children (M Error = 4.36).

Influence of Postdictions on Predictions

Mean postdiction and prediction values across trials for the think-back group are presented in Figure 8. Trial N postdictions were significantly less than Trial N + 1 predictions on each of the four trials (Trial 1: $t(29) = -3.64$, $p < .05$; Trial 2: $t(32) = -4.24$, $p < .001$; Trial 3: $t(31) = -5.58$, $p < .001$; Trial 4: $t(32) = -4.93$, $p < .001$). Thus, even though most children's postdictions tended to be accurate, their predictions appear not to have been influenced by their postdictions. Children's overwhelming tendency was to

Figure 8. Mean postdictions and predictions across trials for the think-back group of Experiment 3.

predict that they would recall more items on an upcoming trial than they remembered having recalled on the previous trial.

Relationships between Age, Self-Predictions, and Recall

Correlations among the self prediction values, recall values, and age are presented in Table 3. As in Experiments 1 and 2, stable individual differences were evident in self-predictions (median cross-trial r was .52) and to a lesser degree, in actual recall (median cross-trial r was .35). However, with one exception, the number of pictures that children predicted they would recall on a particular trial was not correlated with the number they actually recalled on the previous trial. The exception was a small positive correlation between recall on Trial 3 and prediction for Trial 4 ($r = .25$). Whereas in Experiment 2 recall performance on the fourth trial was significantly correlated with predictions on the fifth trial, this finding was not replicated in the current experiment.

Correlations between recall on Trial N and prediction for Trial $N + 1$ across the three experiments are presented in Table 4. There is no evidence of a relation between prediction and previous trial's recall for either the second or third prediction made. However, a small significant relation is evident for the fourth and fifth prediction.

As in Experiments 1 and 2, self-predictions were not significantly related to age on any trial. However, in a replication of Experiment 2, recall on certain trials was significantly correlated with age. Two correlations between recall on a particular trial and age were significant, and all five were positive (see Table 3), with median $r = .21$. Total recall was also significantly correlated with age, $r(64) = .35, p < .01$. These

Table 3. *Intercorrelations Among Self-predictions, Recall, and Individual Difference Measures in Experiment 3.*

	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Age	-.03	.05	-.04	.08	.16	.42**	.21	.30*	.17	.20	.42*	.48*	.27*
2. Prediction 1		.35*	.26*	.35*	.43**	-.06	-.11	-.00	.02	-.21	.06	-.04	.18
3. Prediction 2			.59**	.50**	.47**	.00	.02	.18	-.08	-.19	.26*	.15	.15
4. Prediction 3				.72**	.67**	.04	.13	.21	-.07	-.03	.22	-.03	.18
5. Prediction 4					.85**	.09	.22	.25*	-.05	-.06	.23	.21	.23
6. Prediction 5						.17	.17	.25*	.07	-.04	.26*	.28*	.21
7. Recall 1							.37**	.44**	.48*	.30*	.19	.25*	.26*
8. Recall 2								.38**	.30*	.33**	-.03	.21	.29*
9. Recall 3									.40*	.24	.13	.29*	.35**
10. Recall 4										.29*	.12	.17	.24
11. Recall 5											.05	-.03	.07
12. Cardinality												.28*	.12
13. Pseudoword													.37**
14. Listening Span													

Note: * $p < .05$; ** $p < .01$

Table 4. *Correlations Between Recall on Trial N and Prediction for Trial N + 1 Over Experiments 1, 2, and 3.*

Recall (Trial) – Prediction (Trial)	Exp. 1	Exp.2	Exp.3	Avg. r
Recall (1) – Prediction (2)	.08	.26	.00	.11
Recall (2) – Prediction (3)	.04	.00	.13	.06
Recall (3) – Prediction (4)		.19	.25*	.22*
Recall (4) – Prediction (5)		.41*	.07	.25*

* $p < .05$

findings replicate Experiment 2 and several other studies that have found older preschoolers to recall more pictures than younger ones (see Perlmutter & Lange, 1978; Schneider & Pressley, 1997). However, it cannot be concluded that over-prediction errors were generally smaller in older preschoolers than in younger ones in the current experiment. As in Experiment 2, even though recall increased with age and the level of recall that children predicted did not, age was not significantly correlated with the difference between self-predicted recall and actual recall in the aggregate ($r = -.08$) or on the second, third, fourth, or fifth trials ($r_s = -.04, -.18, .02, .04$, respectively). However, unlike Experiment 2, age was significantly related to this difference on the first trial ($r(62) = -.25, p < .05$). Thus, overprediction errors were smaller for older children on the first trial. This moderate correlation may suggest that older children are more accurate initially compared to younger children but that all children perform similarly overall. But before drawing this conclusion, this finding should be replicated because the correlation may be due to chance.

Understanding of Cardinality

Twenty-three of the sixty-four children performed at chance or below on the cardinality test whereas forty-one answered at least three of the four test questions on this test correctly (overall $M = .71, SD = .29$), which was significantly greater than chance, $t(62) = 15.69, p < .001$. Understanding of cardinality was significantly correlated with age, $r(62) = .42, p < .001$ (see Table 3). Four-year-olds performed significantly lower than 5-year-olds on this test ($M = .60$ and $.82$, respectively), $t(62) = 3.40, p < .01$. Sixty-four percent of the children passed this task compared to 50% in Experiment 2. The

higher passing rate is likely the result of the age difference between the participants in the two experiments. The mean age in Experiment 3 ($M = 4 - 11$) was three months greater than that in Experiment 2 ($M = 4 - 8$).

Contrary to the findings of Experiment 2, understanding of cardinality was positively correlated with children's predicted recall on a few trials (see Table 3), as well as with their average predicted recall over the five trials combined, $r(62) = .28, p < .05$. The addition of the think-back group to the current experiment is the likely reason for the discrepancy. Understanding of cardinality was positively correlated with average predicted recall in the think-back group, $r(31) = .46, p < .01$, but not in the control group, $r(29) = .03, p > .10$. After reporting how many pictures they had recalled on a trial, those children who had a better-developed appreciation for the relative value of numerals (e.g. understood that "eight" was more than "six" or that "five" was less than "seven") predicted that they would recall more than the other children predicted that they would recall. This result is quite strong evidence against the hypothesis that a poor understanding of cardinality underlies young children's tendency to overpredict their recall. If anything, a developing understanding of cardinality promotes overprediction, although it only promotes it when children are required to think back to their recall on the last trial before predicting recall on the next trial.

Understanding of cardinality was not significantly related to mean recall performance, $r(62) = .12, p > .10$. Therefore, in the think-back group, those who showed a better understanding of cardinality tended to overpredict recall by a larger margin, $r(31) = .38, p < .05$. The average margin for those who answered at least three of the four

cardinality questions correctly was 4.69 ($SD = 2.13$, $N = 24$), compared to 2.59 ($SD = 3.11$, $N = 9$) among the remaining children, $t(31) = 2.22$, $p < .05$.

Whereas the understanding of cardinality was associated with poorer prediction accuracy in the think-back group, its association with postdiction accuracy in this group was in the opposite direction, $r(31) = .32$, $p = .068$. Those children who understood cardinality the best overpredicted their recall by the largest margins even though they were able to recollect past recall performance more accurately than the other children. This finding is further evidence against the monitoring deficiency explanation for children's persistent overconfidence in their recall memory. Additionally, the performance of children who passed the cardinality test and those who did not pass was not significantly different in overall mean recall performance nor overall postdiction accuracy, $ts < 1$, $p > .05$. Lastly, understanding of cardinality was significantly correlated with scores on pseudoword repetition, $r(62) = .28$, $p < .05$.

Working Memory and its Relation to Recall, Prediction, and Postdiction

Two measures of working memory were administered, pseudoword repetition (Gathercole, 1995; Merriman et al., in press) and a complex listening span task (Siegel & Ryan, 1989; Leather & Henry, 1994). Each task assessed a different component of working memory. Pseudoword repetition is a measure of the phonological store and complex listening span is a measure of simultaneous memory storage and processing.

Pseudoword Repetition Performance is summarized in Table 5. A 2 (age: 4- vs. 5-year-olds) x 3 (word length: 2, 3, 4 syllables) mixed analysis of variance of rate of

correct repetition yielded a significant effect of age, $F(1,62) = 4.47, p < .05$. Five-year-olds correctly repeated back a greater proportion of pseudowords than 4-year-olds ($M_s = .72$ and $.79$, respectively). There was also a significant effect of word length, $F(2,62) = 12.78, p < .001$. As word length increased, performance declined. These results are consistent with previous findings (Gathercole, 1995; Merriman & Lipko, 2008; Merriman et al., in press).

Table 5. *Mean Proportion of Correct Pseudoword Repetitions in Experiment 3.*

Age	N	2 Syllable	3 Syllable	4 Syllable	Total
4	31	.84 (.18)	.62 (.18)	.68 (.25)	.72 (.11)
5	33	.87 (.15)	.78 (.18)	.68 (.32)	.79 (.13)

Standard deviations are in parentheses.

Pseudoword repetition was expected to be significantly related to prediction accuracy. Specifically, it was hypothesized that children with higher working memory scores would demonstrate less overconfidence overall and make more calibrated predictions across trials compared to children with lower working memory scores. Contrary to prediction, a child's pseudoword repetition score was not related to the average margin by which he or she overpredicted recall, $r(62) = .03, p > .10$. It was also not related to this margin on any particular trial except Trial 5, $r(62) = .25, p < .05$. This last finding is in the opposite of the predicted direction. On average, those children who repeated back the most pseudowords tended to recall more pictures than other children, $r(61) = .25, p < .05$, but also showed a non-significant tendency to predict that they would

recall more pictures, $r(62) = .15$, $p > .10$ (See Table 3 for relations to recall and prediction on specific trials). Because these two tendencies offset each other, pseudoword repetition was unrelated to prediction accuracy. When the analyses just reported were conducted for the think-back and control groups separately, results were quite similar.

Although not the focus of the current study, pseudoword repetition was found to be related to children's understanding of cardinality, $r(62) = .28$, $p < .05$, but not their overall postdiction accuracy, $r(62) = -.01$, $p > .10$.

Complex Listening Span. Children's average listening span score was 1.78 ($SD = .55$, Range = 1 to 3.42, Max Possible = 5). Replicating Leather and Henry (1994), span scores tended to increase with age, $r(61) = .28$, $p < .05$.

Again, children with higher working memory scores were expected to show less overconfidence overall and to make more calibrated predictions across trials compared to those children with lower working memory scores. Contrary to prediction, listening span was not related to the average margin by which a child overpredicted recall, $r(61) = .08$, $p > .10$. It was also not related to the margin by which a child overpredicted recall on any particular trial, r 's ranged from .00 to .15, all $p > .10$. On average, those children with the higher span scores tended to recall more pictures than other children, $r(61) = .34$, $p < .01$, but also showed a non-significant tendency to predict that they would recall more pictures, $r(61) = .24$, $p = .063$ (see Table 3 for relations to recall and prediction on specific trials). Because these two tendencies offset each other, listening span was unrelated to prediction accuracy. When the analyses just reported were conducted for the think-back and control groups separately, results were quite similar.

Although not the focus of the current experiment, listening span was not found to be related to either children's understanding of cardinality, $r(61) = .13, p > .10$, or their overall postdiction accuracy, $r(61) = .22, p > .10$.

Both listening span and pseudoword repetition were developed to measure different aspects of working memory. Pseudoword repetition assesses children's efficiency in verbal storage whereas listening span assesses simultaneous memory storage and processing of verbal stimuli. Scores on these tasks were moderately intercorrelated, $r(62) = .37, p < .01$, which is consistent with their working memory interpretation. A measure of global efficiency of working memory was created by adding together z-scores on the two working memory tasks. As was the case for the individual measures, the global measure was not related to the average margin by which a child overpredicted recall, $r(61) = .07, p > .10$. Those children with higher scores tended to recall more pictures than other children on average, $r(61) = .36, p < .01$, but also showed a non-significant tendency to predict that they would recall more pictures, $r(61) = .23, p = .069$. Because these two tendencies offset each other, the global scores was unrelated to prediction accuracy.

CHAPTER 5

GENERAL DISCUSSION

Summary of Main Findings and Implications

Across three experiments, young children overconfidently predicted their picture recall. This overconfidence may seem surprising, especially given that their overconfidence persisted after a great deal of task experience in which children performed substantially lower than they had predicted. Additionally, this overconfidence may be surprising because children made their predictions after studying the pictures which remained in view at the time of the prediction.

Previously, the wishful thinking hypothesis (Stipek et al., 1984; Schneider, 1998) has been offered as an explanation for such overconfidence. According to this hypothesis, children base their predictions on their desires for performance rather than their actual expectations. Thus, the hypothesis assumes that children desire higher levels of performance for themselves and as a result, children predict their own performance more overconfidently than they predict another child's performance. However in Experiment 2, children's self and other predictions on the picture recall task were equally overconfident. Thus, children's desires to perform better than other children on the present memory task cannot account for their persistently overconfident predictions.

An alternative explanation for children's persistent overconfidence, performance monitoring failures, was evaluated in Experiment 3. Specifically, one group (think-back) of children was asked to make a postdiction for their previous trial's recall before making a prediction for their next trial's recall. Consistent with previous research (Schneider, 1998), children made fairly accurate postdictions suggesting that they can accurately monitor their performance. However, children's predictions in the two groups (think-back vs. control) did not differ. Thus, the children who made postdictions prior to making their predictions were not less overconfident across trials compared to the children who did not make postdictions. Even though these postdictions may have caused a child's own performance on a recall trial to become more salient to him or her, evidence does not exist that the increased salience of such information influenced the child's prediction of upcoming recall performance.

Lastly, the current experiments evaluated children's understanding of cardinality and their working memory capacity as potential explanatory factors for children's persistent overconfidence. Children's overconfidence cannot be explained through a lack of understanding of cardinality. Children who did not understand that "seven" was larger than "five," for example, were no more overconfident across trials compared to those children who had achieved this understanding. Finally, contrary to expectation, prediction accuracy was not related to children's performance on either of two working memory tasks.

*Why Might Children Neglect Past Performance
When Making Predictions?*

One possible reason that preschoolers' predictions are not influenced by previous experience is that young children do not commonly evaluate their own abilities and thus do not have much experience adjusting their expectations based on previous experience. In early education settings, teachers tend to focus on learning goals rather than performance goals. Mastery of skill is encouraged rather than competition among children and awards (e.g. Midgley, Kaplan, & Middleton, 2001). Thus, children may be rewarded for trying and working on a task regardless of their actual performance levels. Although such reinforcement patterns are likely to help maintain and promote motivation and self-esteem, they do not give children experience with true evaluations of their abilities.

Another possible reason is that using previous experience to predict future performance is a learned strategy and preschoolers tend not to spontaneously use strategies (e.g. Kail, 1990; Kruetzer, Leonard, Flavell, & Hage, 1975). In addition to failing to appreciate the benefits of strategies, using a strategy is often costly (in effort) for young children. It can be difficult to learn and correctly use a strategy and in certain situations the costs may outweigh the benefits (Siegler & Alibali, 2005). Additionally, if children are overconfident, it is unlikely that they will be aware that they even need a strategy and thus they will not search for one.

Lastly, it may be adaptive for children to neglect their previous experience. Overconfidence can be adaptive and even beneficial in certain situations (Shin et al., 2007) because it can help to maintain children's motivation levels on a task and

encourage them to try new things (Bandura, 1989). Thus, failing to consider task experience at the time of prediction may be part of this adaptive response. If children are influenced by their past poor performance, they may stop performing a task or lower their level of attention and effort while working on the task. Future studies could examine this possibility by making children's previous failure experience very salient and measuring children's levels of compliance with the task and overall completion rate.

Limitations of the Current Experiments

One limitation of the investigation was that only one method was used to evaluate the wishful thinking hypothesis. In Experiment 2, this hypothesis was evaluated by comparing children's self and other predictions for the picture recall task. The hypothesis assumes that children wish to perform better than others on a task, and so base their self predictions on their wishes and their other predictions on true expectations of performance. However, it is possible that children are not competitively motivated on such tasks and thus do not desire their performance to be better than the performance of others.

The wishful thinking hypothesis has also been evaluated by asking children to differentiate between their wishes and expectations. For example, Schneider (1998) asked children to predict how far they wished they could jump from a starting line and how far they expected to jump from a starting line by placing a different color flag at each position. In four tasks (two cognitive and two psychomotor), children did not differentiate between their wishes and expectations, thus supporting the wishful thinking hypothesis. In future studies using the picture recall task, children should be asked to

state how many pictures they wish they could recall and how many pictures they expect to recall on each trial and the two responses should be compared. Without this additional assessment, it is not possible to completely discount the wishful thinking hypothesis as an explanation for children's overconfidence on the picture recall task.

Another limitation of the current experiments was the limited availability of working memory tasks for the preschool age group. Young children's performance levels on some common working memory tasks are at ceiling or very low and thus they must be modified or cannot be used with preschoolers (Gathercole et al., 2004). Thus, the choices for working memory tasks for the current experiments were limited. However, it is unlikely that any working memory task would account for children's overconfidence. Both of the working memory tasks and the understanding of cardinality task are measures of cognitive maturity and thus high scores would be expected to be related to more accurate predictions, yet none of the tasks were related to children's prediction patterns.

Lastly, due to geographic limitations, the sample who participated in the experiments was not racially or ethnically diverse. This lack of diversity may limit the generalizability of the current experiment's findings. Future work should examine overconfidence in a more diverse sample.

Future Directions for Research on Children's Overconfidence

Future work should investigate children's confidence in their predictions. How confident or certain children are in their predictions across trials may provide a more detailed picture of how children make performance predictions. Assigning a low

confidence rating to a prediction could signal that children are guessing or are aware that their predictions are overconfident. Asking children to make confidence ratings could also directly influence children's recall predictions. If children realize that they assigned a high confidence rating to a prediction of nine pictures yet only recalled four, this discrepancy may influence their later predictions by highlighting previous performance failure.

Researchers should also investigate the role of perceived effort in children's predictions. Young children have a tendency to equate high effort levels with high performance levels (Stipek & Mac Iver, 1989). Thus, their predictions may be based on how hard they intend to try on each trial rather than on how many pictures they think they can actually recall. If children's predictions are influenced by perceived effort, higher effort ratings should be correlated with higher predictions. Similar to confidence ratings, effort ratings could directly influence children's predictions. If children provide a high effort rating and a high prediction for a trial and their performance is lower than their prediction, the discrepancy may increase the salience of their performance and thus influence their next prediction. Additionally, asking children to make effort ratings *after* they have predicted and attempted recall may serve to highlight a mismatch between effort and performance. For a child who overpredicts her recall and then states that she tried really hard on that trial, the realization that her high effort level did not lead to optimal performance should be quite salient or at least could be experimentally manipulated to be very salient.

In Experiment 3, asking children how many pictures they remembered on a just-completed recall trial did not cause them to lower their prediction for the next trial.

Future research should examine how children think about the discrepancy between their rather poor immediate past performance and the rather good performance that they anticipate in the immediate future. Failure on a task can be attributed to either lack of ability (helpless orientation) or to the task itself (mastery orientation). Such attribution styles influence children's levels of motivation and expectations for future success (e.g. Elliott & Dweck, 1988).

Additionally, attribution styles can be influenced by individuals' task goals. For example, children who have performance goals tend to focus on overall outcome levels rather than on what can be learned from the experience of performing the task (Midgley et al., 2001; Elliott & Dweck, 1988). Such goals can be adaptive and have been related to persistence on a task (Midgley et al., 2001). Thus, investigating the types of goals children set for themselves on tasks may help to further explain children's overconfidence in the face of repeated failure.

Another future direction is to have children predict how many pictures they will forget rather than how many they will recall. Predicting one's errors may involve different processes than predicting one's success. For example, Koriat, Bjork, Sheffer, & Bar (2004) demonstrated that predicting errors influenced participants to more heavily consider the role of retention intervals in forgetting when making their predictions. Data patterns in which children provided persistently low error predictions across trials would suggest that children are truly unrealistic in their assessment of their own performance.

Such converging evidence would suggest that regardless of the type of prediction (error vs. success) preschoolers cannot accurately assess their performance on a task.

General Conclusions

Overconfidence declines with age (Schneider & Presseley, 1997), yet older children and adults remain overconfident in certain situations (e.g. Metcalfe, 1998; Plumert, 1995; Plumert & Schwebel, 1997). Because overconfidence can be detrimental in circumstances like educational settings in which accurate evaluation of learning is crucial, it is important to understand the mechanisms of overconfidence at all ages. The results of the current experiments pose a challenge for three potential explanations for preschoolers' persistent overconfidence namely, wishful thinking, memory monitoring limitations, and understanding of cardinality. Although, many factors are likely to influence the accuracy of children's performance predictions, children's abilities to and motivation to use previous experience to guide their predictions of future performance is one factor in need of further investigation.

REFERENCES

- Alloway, T.P., Gathercole, S.E., Willis, C., & Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, 87, 85-106.
- Baddeley, A.D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.
- Baddeley, A.D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417-422.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology*, 25, 729-735.
- Bjorklund, D.F., & Green, B.L. (1992). The adaptive nature of cognitive immaturity. *American Psychologist*, 47(1), 46-54.
- Bjorklund, D.F., Coyle, T.R., & Gaultney, J.F. (1992). Developmental differences in the acquisition of an organizational strategy: Evidence for the utilization deficiency hypothesis. *Journal of Experimental Child Psychology*, 54, 434-448.

- Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, 61 (12, Serial No. 246).
- Cunningham, J.G., & Weaver, S.L. (1989). Young children's knowledge of their memory span: Effects of task and experience. *Journal of Experimental Child Psychology*, 48, 32-44.
- Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Dunlosky, J. & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging*, 15(3), 462-474.
- Elliott, E.S., & Dweck, C.S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54(1), 5-12.
- Flavell, J.H., Friedrichs, A.G., & Hoyt, J.D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, 1, 324-340.
- Gathercole, S.E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition*, 23, 83-94.
- Gathercole, S.E., Pickering, S.J., Ambridge, B., Wearing, H. (2004). The structure of working memory from 4 to 15 years age. *Developmental Psychology*, 40(2), 177-190.

- Henry, L.A., & Norman, T. (1996). The relationships between memory performance, use of simple memory strategies, and metamemory in young children. *International Journal of Behavioral Development, 19*(1), 177-199.
- Hertzog, C., Dixon, R.A., & Hultsch, D.F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging, 5*, 215-227.
- Hertzog, C., Saylor, L.L., Fleece, A.M., & Dixon, R.A. (1994). Metamemory and aging: Relations between predicted, actual, and perceived memory task performance. *Aging and Cognition, 1*, 203-237.
- Kail, R. (2002). Developmental change in proactive interference. *Child Development, 73*, 1703-1714.
- Kail, R. (1990). *The development of memory in children* (3rd ed.). New York: W.H. Freeman and Company.
- Koriat, A., Bjork, R.A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology – General, 133*(4), 643-656.
- Kreutzer, M.A., Leonard, C., Flavell, J.H., & Hagen, J.W. (1975). An interview study of children's knowledge about memory. *Monographs of the Society for Research in Child Development, 40*(1), 1-60.
- Leather, C.V., & Henry, L.A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology, 58*, 88-111.

- Lipko, A.R., Dunlosky, J., Rawson, K.A., Swan, K., & Cook, D. (2007). Using feedback to improve undergraduate and middle-school students' metacomprehension accuracy. Submitted Manuscript.
- Merriman, W.E., & Lipko, A.R. (2008). A dual criterion account of the development of linguistic judgment in early childhood. *Journal of Memory and Language*, 58, 1012-1031.
- Merriman, W.E., Lipko, A.R., & Evey, J.A. (in press). How young children judge whether a word is one they know: A dual criterion account. *Journal of Experimental Child Psychology*.
- Metcalfe, J. (1998). Cognitive Optimism: Self-deception or memory-based processing heuristics. *Personality and Social Psychology Review*, 2(2), 100-110.
- Midgley, C., Kaplan, A., Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology*, 93(1), 77-86.
- Okamoto, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development*, 61, 27-58.
- Perlmutter, M., & Lange, G. A. (1978). A developmental analysis of recall-recognition distinctions. In P. A. Ornstein (Ed.), *Memory development in children*. Hillsdale, NJ: Erlbaum.
- Plumert, J.M., Schwebel, D.C. (1997). Social and temperamental influences on

children's overestimation of their physical abilities: Links to accidental injuries.

Journal of Experimental Child Psychology, 67, 317-337.

Plumert, J.M. (1995). Relations between children's overestimation of their physical abilities and accident proneness. *Developmental Psychology*, 31(5), 866-876.

Pressley, M., & Ghatala, E. S. (1989). Metacognitive benefits of taking a test for children and young adolescents. *Journal of Experimental Child Psychology*, 47, 430-450.

Pressley, M., Ross, K. A., Levin, J. R., & Ghatala, E. S. (1984). The role of strategy utility knowledge in children's strategy decision making. *Journal of Experimental Child Psychology*, 38, 491-504.

Pressley, M., Levin, J. R., & Ghatala, E.S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning and Verbal Behavior*, 23, 270 – 288.

Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition, and wishful thinking. *Developmental Science*, 1(2), 291-297.

Schneider, W., & Lockl, K. (in press). Procedural metacognition in children: Evidence for developmental trends. In J. Dunlosky, & R.A. Bjork (Eds.) *Handbook of Metamemory and Memory*. (pp. 391-409) Taylor & Francis.

Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (second edition). Mahwah, NJ: Erlbaum, 1997.

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

- Shin, H., Bjorklund, D.F., Beck, E.F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development*, 22, 197-212.
- Siegel, L.S. & Ryan, E.B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, 60, 973-980.
- Siegler, R.S., & Alibali, M.W. (2005). *Children's Thinking*. (4th edition). New Jersey: Prentice Hall.
- Stipek, D.J., & Hoffman, J.M. (1980). Development of children's performance-related judgments. *Child Development*, 51(3), 912-914.
- Stipek, D.J., Roberts, T.A., & Sanborn, M.E. (1984). Preschool-age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. *Child Development*, 55, 1983-1989.
- Thiede, K. W. (1999). The importance of monitoring and self-regulation during multi-trial learning. *Psychonomic Bulletin & Review*, 6, 662-667.
- Towse, J. N., & Hitch, G. J. (2007). Variation in working memory due to normal development. In A.R.A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in Working Memory* (pp. 109-133). New York: Oxford University Press.
- Yussen, S.R., & Levy, V.M. (1975). Developmental changes in predicting one's own span of short-term memory. *Journal of Experimental Child Psychology*, 19, 502-508.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence* (3rd ed.).

The Psychological Corporation.

Wilkes, A. (1999). *My first word book*. New York: DK Publishing, Inc.

APPENDIX A

PICTURES AND PRESENTATION ORDER USED IN PREDICTION TASKS

Experiment 1 Stimuli

Trial 1: Blocks, Bike, Plant, Horse, Chair, Lion, Teddy-bear, Car, Ball, Sneakers

Trial 2: Keys, Shirt, Shovel, Bird, Ice Cream, Cat, Truck, Doll, Butterfly, Shark

Trial 3: Bucket, Snake, Toothbrush, Sunglasses, Coat, Boat, Dog, Grapes, Clock, Cup

Experiments 2 & 3 Stimuli

Trial 1: Cat, Juice, Car, Sneakers, Cookies, Keys, Teddy-bear, Crayons, Bird, Ice Cream

Trial 2: Butterfly, Shirt, Banana, Heart, Toothbrush, Ball, Snake, Light-bulb, Shark, Pig

Trial 3: Gloves, Carriage, Dress, Truck, Balloon, Shovel, Doll, Bucket, Horse, Sunglasses

Trial 4: Lion, Blocks, Grapes, Boat, Rocks, Train, Cup, Clock, Bug, Cheeries

Trial 5: Dog, Paper, Bike, Chair, Duck, House, Fish, Plant, Coat, Leaves

APPENDIX B

WORDS AND PRESENTATION ORDER FOR PSEUDOWORD REPETITION TASK (EXPERIMENT 3)

Practice Word:	Chair	
Test Words:	Arm	Driver
	Hate	Brastering
	Pot	Bird
	Bird	Dopelate
	Roobid	Kannifer
	Button	Rabbit
	Grundle	Parrazon
	Letter	Commeecitate
	Diller	Stoppogratic
	Penned	Confrantual

APPENDIX C

SENTENCES, EXPECTED WORDS, AND PRESENTATION ORDER USED FOR LISTENING COMPLEX SPAN TASK (EXPERIMENT 3)

Practice Set:

When I have a cold, I blow my ____ (nose)

When I hear a funny joke I ____ (laugh)

Set 2:

The color of grass is ____ (green)

I go to the doctors when I am ____ (sick)

Set 2:

I can hear with my ____ (ears)

The color of the sky is ____ (blue)

Set 2:

I wear a hat on my ____ (head)

You drink coffee out of a ____ (cup/mug)

Set 3:

I use an umbrella when it ____ (rains)

When I'm tired I go to ____ (sleep/bed)

I wear shoes on my ____ (feet)

Set 3:

Trees are covered with green ____ (leaves)

I can see the sun up in the ____ (sky)

I wear a scarf around my ____ (neck)

Set 3:

I go swimming in the ____ (pool/ocean/lake)

I eat food with a spoon and ____ (fork/knife)

I can tell the time by looking at my ____ (clock/watch)

Set 4:

When babies are hungry or wet they ____ (cry)

I wear gloves on my ____ (hands)

There are swings and slides at the ____ (park)

I like the game hide and ____ (seek)

Set 4:

In the summer it is very ____ (hot)

People go to see monkeys at the ____ (zoo)

On my toast, I like to spread ____ (butter/jelly)

I use a brush to brush my ____ (hair/teeth)

Set 4:

At night I go to sleep in my ____ (bed)

I eat cereal out of a ____ (bowl)

When it is dark I a room, we turn on the ____ (light)

I can see with my ____ (eyes)

Set 5:

The moon comes out at ____ (night)

Worms live in the ____ (ground)

An animal that “barks” is called a ____ (dog)

Before I go outside in the cold, I have to put on my ____ (coat)

When I eat at the table, I sit on a ____ (chair)

Set 5:

When I want to watch a show, I turn on the ____ (tv)

You go sailing in a ____ (boat)

In the morning, I like to drink ____ (juice)

An animal that “moos” is called a ____ (cow)

At the beach, I make castles out of ____ (sand)

Set 5:

At the library, you can read a ____ (book)

At my birthday party, we had yummy birthday ____ (cake)

An animal that “meows” is called a ____ (cat)

The color of a school bus is ____ (yellow)

When you want to talk to someone, you can call them on the ____ (phone)