Spring 5-25-2016

The Effects of Self-Assisted Monitoring on Children's Recall Predictions

Gina R. Stephan
The College at Brockport, gstep1@brockport.edu

The Effects of Self-assisted Monitoring on Children's Recall Predictions

Gina Stephan

The College at Brockport, State University of New York

Abstract

Young children consistently overestimate their judgments of how well they will perform on a picture recall task compared to their actual performance (Flavell, Friedrichs, & Hoyt, 1970; Lipko-Speed, 2013; Lipko, Dunlosky, & Merriman, 2009). Previous researchers have investigated ways to make children more aware of their actual abilities (Lipko-Speed, 2013; Schneider, 1998; Stipek, Roberts, & Sanborn, 1984). This study examines the influence of self-assisted monitoring on young children's overconfidence. Specifically, children will monitor their own performance on a recall task with the help of an experimenter. Such monitoring is expected to lower children's overconfidence in their future performance predictions on a recall task. 50 four- and five- year olds were randomly matched by gender to one of two groups: an experimenter monitored group or a self-assisted monitored group. All children participated in four trials of a picture recall task during which they made 2 recall predictions and 2 recall attempts per trial, each with different sets of pictures. The procedure for the experimenter monitored group was modeled after Lipko-Speed (2013). Specifically, after each recall attempt, children were told by the experimenter, who had been monitoring their recall, how many pictures they had correctly recalled. In the self-assisted monitored group, children (with some assistance) monitored the accuracy of their own recall attempts. Both groups lowered their predictions within and between trials, however their overconfidence persisted. Children's overconfident performance predictions did not decrease within or between trials in either group. Hence, the implementation of this investigation's self-assisted monitoring task did not aid in decreasing children's overconfident judgment predictions on future tasks.

The Effects of Self-monitoring on Children's Recall Predictions

Previous research has shown that young children are persistently overconfident in their judgments of how they think they will perform, compared to how they actually perform (e.g., Flavell, Friedrichs, & Hoyt, 1970; Lipko-Speed, 2013; Lipko, Dunlosky, & Merriman, 2009; Powel, Bolich, & Stewart, 1993; Powell, Morelli, & Nusbaum, 1994; Schneider, 1998; Shin, Bjorklund, & Beck, 2007; Stipek, Roberts, & Sanborn, 1984; Stipek & Maclver, 1989). Such overconfident judgment predictions may result in negative learning outcomes for children in educational settings. If children believe they have learned material successfully, but have actually overestimated how much they learned, it could result in poor future performance (Lipko-Speed, 2013). With age, overconfidence decreases and awareness of one's own abilities increases. (Serra & Dunlosky, 2005). Although older children are able to make more accurate prediction judgments compared to younger children, their initial predictions are typically above their actual performance level (Clifford, 1978; Flavell et al., 1970; Lipko et al., 2009; Lipko-Speed, 2013; Powell et al., 1993; Powell et al., 1994; Schneider, 1998; Stipek et al., 1984; Yussen & Levy, 1975). Researchers have investigated this persistent overconfidence using many tasks and across a range of ages.

Flavell et al. (1970) were among the first researchers to study children's overconfidence on a memory task. Children in nursery school, kindergarten, second grade, and fourth grade were asked to give predictions of how many pictures they would be able to remember. Children were asked to study ten pictures and then to predict how many out of the ten they would be able to remember. Their prediction memory span was determined by the amount of the longest series of pictures they estimated to remember. After each trial, another picture was added to the original list and children were asked if they would be able to remember the pictures. Across all ages, the children's prediction memory span was higher than their actual memory span. The children in nursery school and kindergarten, however, gave higher predictions of their memory span than second and fourth graders. Thus, younger children were more overconfident when making their predictions, thinking they would be able to remember more than they actually could. Older children displayed some overconfidence but gave overall lower prediction judgments than the younger children. This study prompted future investigations on children's overconfidence in their performance on both memory and psychomotor tasks as well as investigations about reasons why children are so overconfident.

One plausible explanation for children's overconfident performance judgments on physical tasks is the Wishful Thinking Hypothesis, which suggests that young children do not separate their expectations from their wishes. In other words, they expect to achieve what they wish to achieve and as a result, they are overconfident (Stipek et al., 1984). Historically, investigators have examined the Wishful Thinking Hypothesis by asking preschool children to perform psychomotor tasks. Stipek et al. (1984) investigated whether children differentiated their wishes from their expectations. Four-year-olds completed a tower task that required them to pull a string that lifts a cart up the tower. In the cart, there was a marble that had a magnet on it. The researchers were able to either turn the magnet on or off, determining whether the ball stayed in or fell out of the cart. A colored track was inserted on the tower after each trial to indicate where the ball had fallen off. Prior to the start of each trial, children were either asked, "How high do you want the car to go when you play the game?" or, "How high do expect the car to go when you play the game?" They were given five opportunities to see how high they would be able to get the cart to the top of the tower. Before each prediction the experimenter directed the children to the colored track and verbally reminded them where the marble had last fallen off.

They then made a new prediction of how high they could pull the cart up the tower without the marble falling off. Researchers found that children's wishes did not differ from their expectancies. They wished they would be able to pull the cart just as high as they expected to, which supports the Wishful Thinking Hypothesis.

Schneider (1998) further investigated the Wishful Thinking Hypothesis to see if findings from a physical task would extend to memory tasks. Four-, five-, and six-year-old children conducted two experiments involving two psychomotor tasks (jumping task and ball throwing task) and two memory tasks (memory span prediction task and hide and seek task). Each experiment had two conditions, wish and expectation. In the wish condition, children were told to either place a green flag where they wished to jump (jumping task) or put the amount of balls they wished to throw in a different container from which they would be tested with (ball throwing condition). Over three trials, children estimated their performance and were given feedback afterwards. After the last trial, children were asked to move a flag to the place that they had last reached, or place the same amount of balls that they had thrown into the container during the test, into the plastic bucket. Schneider (1998) also investigated the Wishful Thinking Hypothesis by asking children to make a prediction performance for another child performing the same physical task. Each child would watch and make a prediction on how well another child would perform a task, and then they would participate in the same task themselves. Before their participation, they would make a prediction judgment about their performance ability for the task. Studies show that children report that they would perform better on a psychomotor task than another child would; meaning they gave higher performance predictions for themselves than they did for predicting another child's ability (Schneider, 1998). This suggests that children are better able to assess another child's physical ability more than their own; children believe that they had greater performance ability than another child. Such findings regarding the Wishful Thinking Hypothesis provide important insight for children's overconfident judgment predictions. However, the tasks used in the previous studies were psychomotor in nature rather than metacognitive. Thus, children are reflecting on their physical abilities, which may be fundamentally different than assessing their cognitive capabilities. To address this concern, Schneider (1998) also used memory tasks to investigate the Wishful Thinking Hypothesis.

In experiment two of Schneider's (1998) study, children participated in two cognitive tasks: a memory span prediction task and a hide and seek task. In the memory span task, children listened to ten words and were asked to repeat them back to the researcher once they thought they could remember them. In the hide and seek task, children were shown 10 pairs of pictures in which one picture out of a pair was hidden in a house. Children were instructed to put the twin of that picture where its pair was hidden. Across both of these tasks, children were asked to make a wish prediction and an expectation prediction for their performance. Thus, children were asked how many they wished to remember and how many they expected to remember on each trial. Results showed that for both memory tasks, children gave similar values for wish and expectation predictions. Overall, children overestimated their cognitive performance but six-year-olds were more accurate than the four-year-olds in predicting their performance on the task. Schneider's (1998) results, coupled with Stipek et al.'s (1984) findings, indicate that children do not differentiate between their expectations and desires during both physical and cognitive tasks. These studies support the Wishful Thinking Hypothesis which states that children may be establishing their predictions based on what they want to happen and not on what they think they can truly accomplish. However, the literature is mixed regarding this explanation for young children's overconfident performance predictions. For example, Lipko et al.'s (2009) findings with a different task did not support the Wishful Thinking Hypothesis.

Lipko et al. (2009) extended on Flavell's (1970) research and hypothesized that if children were familiar with the task, they could more accurately assess their performance ability and in turn overconfidence would decrease across trials. Investigators implemented a study opportunity prior to children's predictions to see if it had an effect on overconfidence in four-and five-year-old children. In the first experiment, children were shown ten pictures and given ten seconds to study them. Next, experimenters asked the children how many of the pictures they thought they would be able to remember if they were covered. They were then given one minute to recall as many pictures as they could and after the time limit was up the researchers told the children how many pictures they had recalled. The procedure was repeated two additional times where children were presented with a new set of ten pictures each time. Analyses showed that children's overconfidence remained high across trials, however predictions decreased when presented with a new set of pictures. In a second experiment, two trials were added, for a total of five. Lipko et al. (2009) predicted that having five trials of experience with the picture recall task would decrease children's overconfidence in their performance on the task. In Experiment two, four- and five- year-olds made performance predictions for another child as well as for themselves on a picture recall task in order to assess the Wishful Thinking Hypothesis. The results indicated that children's self and other-predictions were not significantly different from each other; children predicted that other children would perform just as well as they would in the memory task. These findings are not consistent with previous research using psychomotor and cognitive tasks that found that children gave more accurate predictions of their own performance than for the performance of others (Schneider, 1999; Stipek et al., 1984). This inconsistency suggests that the Wishful Thinking Hypothesis cannot be the only explanation for young children's persistent overconfidence.

Memory monitoring deficiencies may be another possible explanation for young children's overconfidence (Lipko et al., 2009; Lipko-Speed, 2013; Schneider, 1998). That is, if children cannot accurately monitor and remember their own performance on a task, they will not be able to use that knowledge to predict how they will perform on a task in the future or on similar tasks. Lipko et al. (2009) examined this memory monitoring deficiency hypothesis in her third experiment. Four- and five-year-olds were asked to study ten pictures. After studying, they were asked to predict how many pictures they would be able to remember once they were covered up. Half of the children were asked to make a postdiction before their prediction on the second trial. Specifically children were asked, "How many pictures did you remember last time?" This was repeated across two more trials with ten different pictures on each trial. The results of this experiment showed that children were accurately able to remember the amount of pictures they had recalled in the previous trial. These results are consistent with Schneider (1998), which also found that children's postdictions were relatively accurate. These findings (Lipko et al., 2009; Schneider, 1998) contradict the claim that children have memory monitoring deficiencies; children in fact, seem to be quite accurate when prompted by the researcher to recall their previous performance. Although children were able to accurately remember how they performed on a task, they failed to apply it to their prediction of future performance on a similar task (Lipko et al., 2009; Schneider, 1998) where their performance predictions remained overconfident on subsequent trials.

Lipko-Speed (2013) further investigated the findings of Lipko et al. (2009), specifically the finding that young children's persistent overconfidence remained despite their accurate monitoring ability. Lipko-Speed (2013) asked children to make an additional recall prediction for the same set of pictures immediately after their first judgment prediction and recall attempt of the pictures. More specifically, four-and five-year-olds were presented with and given ten seconds to study ten pictures. They were then asked to predict how many pictures they would be able to remember once they were covered up. Children were given sixty seconds to recall the pictures once covered and then the experimenter told the children how many pictures they actually recalled. Next, the experimenter said, “I am going to ask you to try to remember the pictures hiding underneath the paper again. How many of these pictures will you be able to remember this time?” The children then provided a second recall prediction, recalled as many pictures as they could, and again the experimenter provided feedback. This procedure was then repeated with new pictures for three more trials. It was hypothesized that children would use the feedback given by the experimenters after the first task and apply it to a future task. The results demonstrated that children’s second predictions within each of the four trials were significantly less overconfident than the first predictions for the same set of pictures. This result suggests that children are able to remember how well they performed (after being told by an experimenter) and then apply that knowledge to their second recall predictions for the same set of pictures. However, between trials when new pictures were presented, children’s overconfidence returned. Children were able to adjust and lower their recall prediction when asked to recall the same set of pictures a second time, but when presented with a new set of pictures and asked to study and make a prediction, they were just as overconfident as their very first recall prediction. Children fail to apply their knowledge between trials but appear to be able to do so within trials. This study provides important implications that children can use that knowledge to predict their performance on a future task under certain conditions.

Influence of Feedback on young children's overconfidence

Further investigations regarding why children are persistently overconfident have also explored how children can be more accurate predictors of their performance ability (Clifford,1978; Lipko et al. 2009; Lipko-Speed, 2013; Schneider, 1998; Stipek et al. 1984). Explorations of ways to help children adjust their performance predictions to be a more accurate estimator of their ability have primarily focused on the use of feedback; physically or vocally directing children to their previous performance on a task. Feedback regarding children's previous performance was presented to see if they would apply it to a subsequent prediction on a similar task. It was thought that if researchers directed children to their past performance on a previous task, it would increase the children's awareness of their actual ability. In turn, children would lower their future performance predictions to a more accurate estimate of their skill (Lipko-Speed, 2013; Lipko et al., 2009). For example, Clifford (1978) looked at how age and expectancy influenced children's overconfident judgments on a working puzzle task. Children in his study gave more accurate future predictions when researchers monitored and gave feedback about their past performance. First, third, and fifth graders were studied to see whether visual and auditory feedback improved a child's ability to use their past performance for future predictions. Across four trials, each child was given a puzzle to complete in a certain amount of time. In the control group, children were not given any feedback on how many pieces they correctly placed. At the end of the third trial, the children were asked, "How many pieces do you think you will place correctly before I tell you to stop?" In the feedback group children were given a recording sheet in which the researcher marked off a square for every piece the child had correctly placed in the puzzle. After each trial, the experimenter counted aloud the amount of pieces the child correctly placed. The number total was then placed on the top of the recording sheet. Again, after the third trial, children were asked, "How many pieces do you think you will place correctly before I tell you to stop?" and that number was also recorded on the recording sheet. The results showed that overall children were overconfident but older children were more accurate predictors of their future performance than younger children. The control group gave less accurate (more overconfident) predictions of their future performance than those in the feedback group. More specifically, when researchers overtly monitored how many puzzle pieces the children in the feedback group had correctly placed, children were more accurate in predicting how many pieces they had correctly placed in the fourth trial than those in the control group. These findings suggest that children were able to adjust their performance predictions on a future task based on feedback provided by an experimenter. Children, however, were not asked to give performance predictions across trials, but only before the last trial, so there is no way of knowing if their decrease in overconfident judgments decreased across trials.

Stipek et al. (1984) extended Clifford's (1978) findings regarding how feedback can decrease young children's overconfidence on a psychomotor task. Stipek et al. (1984) investigated whether four-year-olds could make more realistic performance predictions when given visual and auditory feedback about their past performance on a task. Stipek's (1984) study, described earlier, implies that preschool-age children are capable of applying experimenter-provided-feedback about their past performance to judgments about future performance. This is consistent with Lipko-Speed (2013) research. However, the results of Stipek et al.'s (1984) study may have been task dependent in that the children had no control over their performance in the game that they participated. That is, unbeknownst to them, the children were not basing their predictions on their own ability but instead on the experimenter manipulated outcomes. It is possible that their predictions would have been more accurate if their performance reflected their actual skills. Despite this limitation, results suggest that feedback indeed helps children in making more accurate predictions of their performance on psychomotor tasks. Other researchers have attempted to apply these findings to cognitive tasks.

As mentioned earlier, Schneider (1998) also found that children had the ability to accurately relay how well they performed in a task when researchers gave feedback and directed them to their past performance. Lipko et al. (2009) also found that although children were able to accurately recall how well they had just performed on a memory task, they were unable to subsequently apply that knowledge to their future predictions for performance on a similar task. Similarly, Lipko-Speed (2013) demonstrated that after receiving performance feedback and being directed to the fact they would be recalling the same set of pictures as they did in the previous test, children lowered their second prediction on a picture recall task. These findings suggest that there may be ways to decrease children's overconfidence.

External vs Self Monitoring

The current study seeks to further investigate possible ways to reduce children's overconfidence on cognitive tasks. Specifically, this study investigates the influence of a modified self-recording activity on children's performance predictions on a picture recall task. Will participating in a task in which children can visually and physically assess their own performance help to lower their overconfidence when predicting their performance for a future task? Self-recording, or recording of one's own behavior (Broden, Hall, & Mitts, 1971), has been studied to examine whether self-monitoring is more effective than teacher- or external-monitoring in behavior modification (Broden et al., 1971; Dunlap & Dunlap, 1989; Kazdin, 1974; Hallahan, Lloyd, Kneedler, & Marshall, 1982). Based on the existing literature, it is hypothesized that self-assisted monitoring will make a difference in reducing young children's overconfidence on a picture recall task.

Previous research has investigated different types of monitoring and whether the type of monitoring differentially affects children's behavior and learning. Broden et al. (1971) used a single subject design to investigate two different monitoring tasks. Researchers gave two students a self-recording mechanism to see whether it increased study behavior or decreased talk out behavior more than if a teacher was reinforcing their good behavior. Baseline behavior was recorded for seven days, after which the students were each given a recording sheet for the seven following days. Each time the students produced a desirable behavior, they put a plus in a square, and if they executed an undesirable behavior, they put a minus in a square. During the intervention phase, the students also saw a counselor who praised them when they behaved in the desired way. After the self-recording week (intervention phase 1), students were not administered the recording sheets for the following five days (baseline phase 2), and were given them again the following week (intervention phase 2). During baseline phase 2, the students' teachers were instructed to focus on the individual students being observed and whenever they had the chance, to give them praise. Also during this period, another researcher was recording both students' behavior. Results showed that student 1 increased her behavior (studying) during the self-recording phase of intervention phase one. During the first day of baseline phase 2, the student's behavior was still very high and it was noted that the teacher's rate of attention to that student also was high. The second day during baseline phase 2, the student's study behavior declined which shows that she was reliant on the self-recording sheet. This drop in study rate persisted even when the self-recording sheets were administered again during intervention phase 2. Similar results were found with student 2 pertaining to decreasing talk outs. Results suggested that self-recording affected the desired behavior. Self-recording made the students more conscious of their behavior and directed their attention to what they needed to change (i.e. studying or talk outs). However, it is still unclear whether the extent to which external monitoring from the teacher and counselor played a role in the observed behavior changes. Because teacher praise was provided in tandem with self-recording, it is possible that such teacher reactivity may have influenced the implementation of the self-recording activity.

In a related study, Hallahan et al. (1982) examined whether teacher assessment or self-assessment differentially affected the behavior of a child, Peter, with a learning disability. Specifically, the study investigated whether feedback from the teacher compared to self-assessment was more efficacious in increasing his on-task behavior. The self-assessment condition consisted of an audiotape recorder producing a tone every forty-two seconds in which Peter was told to ask himself, “Was I paying attention?”, after which he would record “yes” or “no” on his recording sheet. The teacher assessment condition was very similar except when the tone rang, the teacher told Peter whether he was on task or not and Peter recorded her assessment on his recording sheet. This procedure was done in twenty-minute intervals and the conditions were counterbalanced with intervals of no condition over fifty days. The results suggested that both assessments increased Peter’s on-task behavior but a greater increase was observed when the child did both self-assessment and recording. Even though self-assessment was shown to have a greater impact on Peter’s on-task behavior, the increase was not significantly different from when he received feedback from his teacher. This may have been due to the reactivity between both conditions; Peter knew that he was being observed so he may have been more on task when monitored by his teacher because of the attention he was receiving. This study concludes that self-monitoring is of some importance with regards to changing a target behavior.

Dunlap and Dunlap (1989) also investigated children with learning disabilities in the hopes of providing more insight into the implementation of self-monitoring in behavior modification. Initially, three students were given a subtraction worksheet daily, teachers verbally told the students how to complete it, and then were given feedback on their performance. After this was done for a week, the self-monitoring condition was implemented, the students were instructed in the same way but were given a checklist in which they were to record a plus for each step they did correctly or a minus for each step they did not get right. After, the students handed in their sheets and were given a point for each correctly monitored checklist. The results of this study indicated that when the self-monitoring checklist was implemented, the amount of correct answers for each worksheet across trials increased. They answered more questions correctly when using the self-monitoring checklist compared to when they received teacher feedback. This implies that the students were better able to monitor what steps they were doing wrong and correct them when using the self-monitoring aid. Interestingly, when the self-monitoring checklists were withdrawn the students maintained their improvements and continued to perform with more success than they had before the implementation. This study concludes that self-monitoring checklist's aided the children in responding more accurately and consistently. Based on this research, children have the ability to keep track of their correct and incorrect responses and use them to increase a behavior on future trials. By giving children a physical task to keep track of their mistakes, the children were able to adjust their future performance based on their previous assessment. Another physical recording technique, such as the use of a star chart, has been found to influence children's behavior.

In a case study, Lyon (1983) implemented the use of a star chart to treat a 12-year-old boy's lip-biting habit. The participant had an unpleasant and swollen lip as a result of biting his lip during sleep. The boy's mother had noticed blood on the boy's pillow for two years and his pediatrician referred him to a psychologist to modify his behavior. After the initial interview, the boy started to mark his previous behaviors of lip biting on a star chart. He and his mother were to check his pillow each morning for blood to see if he had been biting his lip during the night. If there was blood on the pillow, the boy was not given a star to mark on the chart. If the pillow was clean, the boy was given a blue star to mark on his chart. If the boy received three blue stars in a row (marked on consecutive days), he was given a gold star, which earned him a reward. By the third week he received stars every night, and over the course of six months, the lip biting stopped completely. In this case, Lyon (1983) implies that the use of self-monitoring with a star chart in combination with providing reinforcements reduced the boy's negative lip-biting behavior.

Current Study

As previously described, the metacognitive literature on overconfident judgment predictions uses the term monitoring to describe the process of an experimenter providing feedback (visual/auditory) to children about their performance to help them use information about their past performance to potentially decrease their predictions about their future performance (e.g. Lipko, 2009; Lipko-Speed, 2013). However, the applied behavioral analysis literature uses the terms self-recording, self-assessment, or self-monitoring when referring to the act of children recording their own performance when attempting to increase or decrease a desired behavior (Broden et al., 1971; Dunlap & Dunlap, 1989; Hallahan, et al., 1982). The current experiment will use the term monitoring to refer to the act of keeping track of a child's performance. Specifically, two types of monitoring will be investigated – self-assisted and experimenter. Experimenter monitoring signifies that the researcher will be monitoring and providing feedback to the child regarding his or her performance. Self-assisted monitoring indicates that children, with the help of the experimenter, will monitor their own performance and provide feedback to the experimenter regarding their own performance ability.

Hypothesis one poses that children's overconfidence will decrease within trials in both the experimenter monitored group and the self-assisted monitored group (replicating Lipko-Speed, 2013). Children will give less overconfident second performance predictions for a set of pictures, compared to their first performance predictions for the same set of pictures. Hypothesis two anticipates that in the self-assisted monitored group, children's overconfidence will decrease between trials. Children will give less overconfident judgment predictions when presented with a new set of pictures. More specifically, it is expected that children's overconfidence on the initial predictions of each subsequent pair of trials will be lower compared to the very first judgment prediction on the initial pair of trials.

Method

Participants

A total of fifty-two children (30 males and 22 females) were recruited from various preschools and children's programs in the surrounding areas of Rochester and Buffalo, New York. Participants were preschool age, between 41 months and 72 months, with a mean age of 57 months (4 years and 9 months) who were all from a diverse socioeconomic background.

Procedure

A letter of informed consent was sent home with the children asking their parents' permission for their child's participation in the study. Before starting the study, children were asked if they wished to participate in the experiment, and upon their assent, the experiment continued. Children received stickers after their participation in the study.

In this between-subjects experimental design, children were matched by random assignment into two groups, counterbalancing gender. One group, of twenty-five participants (11 girls and 14 boys), participated in a procedure similar to Lipko-Speed (2013). This group is referred to as the experimenter monitored group. The second group of 27 participants (11 girls and 16 boys) participated in the same procedure with one important difference. This group is referred to as the self-assisted monitored group. Children were tested individually in a quiet room distant from any distractions. In the experimenter monitored group, children were asked to sit next to the experimenter at a table with a large 3 x 2-foot magnetic board and asked to count to ten. Then, a group of ten 4 x 6-inch colored pictures of objects (see appendix for the list of pictures) was presented face down and the children counted one by one how many pictures there were in the pile. An experimenter then individually placed each picture on the magnetic board. As the experimenter put a picture on the magnetic board the child named each one (replacing any pictures that the child could not name correctly or noting if he/she uses an unusual yet correct label). After all ten pictures were named, children were given ten seconds to study the pictures. The children were then asked, “How many of these pictures do you think you will be able to remember once I cover them up?” A blue towel was then placed over the pictures and the experimenter presented the children with a bowl filled with blocks in the shape of stars. The children were then told, “Ok, I have covered the pictures. Name as many pictures as you can remember. Every time you remember a picture, I will say yes, and I will take a star from the bowl and put it on the table in front of you. If I say no, then you did not remember a picture from the board so I will not take a star out of the bowl (experimenter monitoring).” As children recalled pictures that were underneath the towel, the experimenter said yes to those that were accurate and indicated to the child to take a star out of the bowl and place it in front of them.

Children were given sixty seconds to recall as many pictures as they could. After sixty seconds, the experimenter told the child how many pictures they correctly recalled. The researcher then asked the children to give a recall prediction for the same set of pictures they just attempted to recall. Specifically, the experimenter say, “I am going to ask you to try to remember the pictures hiding underneath the blue towel again. How many of these pictures will you be able to remember this time?” After giving their prediction, children were given sixty seconds to recall as many pictures as possible. Following their second recall attempt, the pictures were uncovered and the experimenter told the children the number of pictures they remembered correctly. This entire procedure was repeated three more times with new sets of 10 pictures for each pair of additional trials. Researchers recorded each child’s recall predictions and performance.

The children in the self-assisted monitored group participated in the same procedure as the experimenter monitored group with one important change. Children in this group were asked to monitor their own recall in combination with the experimenter assisting the children in their monitoring. Specifically, children were given a bowl filled with ten small blocks in the shapes of stars. After the children named, studied, and made a recall prediction, they were presented with the bowl of stars and told, “Every time you remember a picture, I will say yes, and you will take a star from the bowl and put it on the table next to you. If I say no, then you did not remember a picture from the board so do not take a star out of the bowl (self-assisted monitoring).” As children accurately recalled pictures the experimenter said yes, and upon that cue, children took a star out of the bowl and placed it in front of them. After the sixty seconds passed, the children were told, “Ok, time is up! Can you count how many stars are on the table?” After children counted out loud the amount of stars that were placed in front of them, they were then asked, “How many pictures from the board did you remember?” If children said that they did not remember or did not know, the experimenter reminded them, "Remember that every time you remembered a picture, you took a star out of the bowl. How many pictures from the board did you remember?" Once the child gave a self-monitored number, the procedure continued in the exact same fashion as in the experimenter-monitored group for three more trials.

Results

Analyses were first conducted with gender as a factor, but the main effect of gender and all interactions including it were not significant, so it is not discussed further. Two tailed mixed analyses of variance were conducted to test the two proposed hypotheses. The first hypothesis stated was that children's overconfidence would decrease within trials in both the experimenter monitored group and the self-assisted monitored group (replicating Lipko-Speed, 2013). The second proposed that in the self-assisted monitored group, children's overconfidence would decrease between trials.

Each child was scored individually. Mean prediction accuracy (predicted recall - actual recall) was calculated for each trial (Table 1). Children made two predictions and two recall attempts on each trial. In order to test hypothesis 1, a 2 (Task: first prediction or second prediction) x 2 (Group: self-assisted vs. experimenter) x 4 (Trial: 1, 2, 3, or 4) mixed analysis of variance (ANOVA) was conducted of children's prediction accuracy and yielded a significant main effect of task, $F(1,50) = 12.88$, $p < .001$, $n^2 = 0.81$ and a significant interaction between task and trial, $F(3,150) = 2.86$, $p < .001$. No significant main effects were found for group. Prediction accuracy in the experimenter monitored group did not differ from prediction accuracy in the self-assisted monitored group, ($F < 1$). The act of monitoring their own performance did not improve children's prediction accuracy within trials. Thus, the first part of hypothesis 1 was not supported. However, two tailed follow-up *t*-tests revealed that children's second predictions were significantly more accurate than their first predictions on trial 4, $t(51) = 4.44$, $p < .001$, Cohen's $d = 0.62$ and approaching significance on trial 3, $t(51) = 1.96$, $p = .055$. However, children's second predictions were not significantly more accurate than their first predictions on the first two trials, $ts < 1.34$ (Table 2) contrary to the findings of Lipko-Speed (2013) that found that children's prediction accuracy improved within a trial across all four trials. Thus, the second part of hypothesis 1 was only partially supported.

Hypothesis two predicted that children would give less overconfident judgment predictions when presented with a new set of pictures. This hypothesis was not supported. Surprisingly, the finding was in the opposite direction. Children's first predictions in Trial 1 were actually significantly more accurate than their first predictions in Trial 4, $t(51) = -2.52$, $p = 0.015$, $d = .44$. This was also approaching significance between Trials 3 and 4; $t(51) = -2.31$, $p = 0.025$. Thus, children's overconfidence between the first and last trial surprisingly increased rather than decreased or remained the same.

Across trials, children's second judgment accuracy scores were not significantly different than first judgment accuracy scores on the following trial, $ts(51) < -1.05$, $ps > .05$ (Table 3), except between Trial 3 and Trial 4, $t(51) = -3.60$, $p = .001$, $d = 0.50$. Children's second predictions on a trial were not significantly more accurate from their first predictions on the immediately subsequent trial. However, children's second predictions in Trial 3 were significantly more accurate than their first predictions on Trial 4.

<u>Prediction Magnitudes</u>

Pertaining to prediction magnitudes, means for each trial are reported in **Table 1**. The first and second prediction magnitudes were compared across trials using a 2 (Prediction Timing: first prediction or second prediction) x 2 (Group: self-assisted vs. experimenter) x 4 (Trial: 1, 2, 3, or 4) mixed analysis of variance (ANOVA) which yielded a significant main effect of prediction timing, $F(1,50) = 20.71$, $p < .001$, $n^2 = 0.46$ and trial, $F(1,50) = 20.14$, $p < .001$, $n^2 = 0.54$ but no significant interactions.

On each trial, children's first predictions were significantly higher in magnitude than their second predictions, $ts(51) > 2.33$, $ps < .01$, $ds > 0.23$ (Table 7). Children predicted that they could recall more pictures in their first prediction compared to their second prediction for the same set of pictures. Taken together, the data replicates Lipko-Speed (2013), specifically looking at predictions within a trial. Children appeared to have some knowledge that their past performance predicts their future performance within the same set of pictures.

When looking at prediction magnitudes across trials, first predictions between Trials 1 and 2, Trials 2 and 3, and Trials 3 and 4 were not significantly different, $ts(51) < 1.80$ (Table 8). However, between Trials 1 and 4, first predictions were significantly different $t(51) = 3.19$, $p = 0.002$, $ds = 0.44$. When presented with a new set of pictures, children's first prediction on trial 4 were significantly lower than their first prediction on Trial 1. Second predictions were also significantly different between trials, $ts(51) > 2.13$, $ps < .038$, $ds > 0.30$ (Table 9), except between Trial 2 and 3 ($t < 1$). Children's second predictions in the previous trial were significantly higher in value than their future predictions.

Between trials when new sets of pictures were presented, there was some suggestive evidence that children were learning from their previous experiences but predictions did not decrease consistently between sequential trials. Children's second predictions on a trial were not significantly different than their first predictions on the subsequent trial, $ts(51) < -1.67$, $ps > 0.101$. Children's second predictions on each trial were not significantly different than their first prediction on the succeeding trial when presented with a new set of pictures. Thus, children did not apply previous knowledge to future performance in a consistent manner.

Recall Magnitudes

Mean recall magnitudes are reported in **Table 1**. Comparing the first and second recall values across trials, a 2 (Recall Timing: first or second) x 2 (Group: self-assisted vs. experimenter) x 4 (Trial: 1, 2, 3, or 4) mixed analysis of variance (ANOVA) yielded significant main effects of timing, $F(1,50) = 56.3$, $p < .001$, $n^2 = 0.03$ and trial $F(3,150) = 37.61$, $p < .001$, $n^2 = 0.97$ but no significant interactions.

In Trials 1, 2, and 3, children recalled significantly more pictures on their first recall attempt than on their second recall attempt within Trials 1, 2 and 3, *ts*(51) > 1.69, *ps* < .098, *ds* > .23 (Table 4). Children recalled more pictures in their first recall attempt than on their second when asked for the same set of pictures.

Mean number of pictures recalled on the first recall attempt were also significantly different across trials, *ts*(51) > 4.90, *ps* < .05, *ds* > 0.68 (Table 5), except between Trials 2 and 3, $t(51) = .82$, $p = .42$. Children's recall ability decreased between trials. They did not significantly remember as many pictures when presented with a new set of pictures as they did in their first recall attempt in a previous trial. However, the number of pictures that children remembered in their first prediction on Trial 2 did not significantly differ from the number of pictures they recalled in their first recall attempt on Trial 3.

Second recalls were also significantly different across trials, *ts*(51) > 3.18, *ps* < .003, *ds* > 0.44 (Table 6), except between Trials 2 and 3, $t(51) = .24$, $p = .81$. In trials 1 and 4, children remembered significantly less pictures when they recalled pictures a second time for the same set of pictures. Additionally, between consecutive recalls, mean number of pictures recalled on the second recall attempt on Trial 1 was significantly greater than the mean number of pictures recalled on the first recall attempt on the second trial, $t(51) = 3.97$, $p < 0.001$, $d = 0.55$. The same result was found between the second recall attempt on the third trial and the first recall attempt on the fourth trial, $t(51) = 4.28$, $p < 0.001$, $d = 0.40$. Children recalled significantly more pictures on their second recall on Trials 1 and 3 than they did on their first recall attempt on Trials 2 and 4.

Table 1

Mean prediction and recall values across trials.

	Accuracy		Prediction		Recall	
Group	Exp.	SA	Exp.	Exp.	Exp.	SA
Trial 1						
First	2.44 (0.78)	3.44 (0.75)	8.00 (0.56)	7.78 (0.65)	2.44 (0.78)	3.44 (0.75)
Second	2.00 (0.63)	2.74 (0.58)	6.69 (0.49)	6.74 (.58)	2.00 (0.63)	2.74 (0.58)
Trial 2						
First	3.68 (0.68)	3.48 (0.71)	7.20 (0.72)	6.81 (0.62)	3.68 (0.68)	3.48 (0.71)
Second	2.96 (0.76)	2.81 (0.64)	5.84 (0.69)	5.78 (0.58)	2.96 (0.76)	2.81 (0.64)
Trial 3						
First	3.48 (0.67)	3.44 (0.46)	6.64 (0.64)	6.67 (0.64)	3.48 (0.67)	3.44 (0.46)
Second	2.80 (0.79)	2.37 (0.71)	5.68 (0.69)	5.19 (0.68)	2.80 (0.79)	2.37 (0.71)
Trial 4						
First	5.36 (0.71)	3.81 (0.62)	7.00 (0.67)	5.52 (0.62)	5.36 (0.71)	3.81 (0.62)
Second	3.16 (0.64)	1.74 (0.58)	5.20 (0.68)	3.44 (.59	3.16 (0.64)	1.74 (0.58)

* *Standard errors in parentheses*

Table 2

Repeated measure t-test results for judgment accuracy within trials

	M	SEM	*t-test*	*p-value*
Overconfidence 1A Overconfidence 1B	0.58	0.43	1.33	0.189
Overconfidence 2A Overconfidence 2B	0.69	0.51	1.35	0.183
Overconfidence 3A Overconfidence 3B	0.88	0.45	1.97	0.055
Overconfidence 3A Overconfidence 4B	2.13	0.48	4.44**	0.000

*$p < 0.01$, **$p < 0.001$
**M = mean sum for each row*

Table 3

Repeated measure t-test results for judgment accuracy between subsequent trials

	M	SEM	*t-test*	*p-value*
Overconfidence 1B Overconfidence 2A	-1.19	0.49	-2.45	0.018
Overconfidence 2B Overconfidence 3A	-.58	0.54	-1.05	0.297
Overconfidence 3B Overconfidence 4A	-1.98	0.55	-3.60*	0.001

*$p < 0.01$, **$p < 0.001$
**M = mean sum for each row*

Table 4

Repeated measure t-test results for mean recall within trials

	M	SEM	*t-test*	*p-value*
List 1 Recall 1 – List 1 Recall 2	0.46	0.22	2.13	0.038
List 2 Recall 1 – List 2 Recall 2	0.50	0.191	2.61	0.012
List 3 Recall 1 – List 3 Recall 2	0.35	0.21	1.69*	0.098
List 4 Recall 1 – List 4 Recall 2	-0.92	0.23	-0.89	0.396

*$p < 0.01$, **$p < 0.001$
**M = mean sum for each row*

Table 5

Repeated measure t-test results for first mean recalls between trials

	M	SEM	*t-test*	*p-value*
List 1 Recall 1 – List 2 Recall 1	1.5	0.31	4.90**	0.000
List 2 Recall 1 – List 3 Recall 1	0.23	0.28	0.822	0.415
List 3 Recall 1 – List 4 Recall 1	1.52	0.27	5.66**	0.000
List 4 Recall 1 – List 1 Recall 1	-3.25	0.33	-9.73**	0.000

*$p < 0.01$, **$p < 0.001$
**M = mean sum for each row*

Table 6

Repeated measure t-test results for mean recall within trials

	M	SEM	*t-test*	*p-value*
List 1 Recall 2 – List 2 Recall 2	1.54	3.27	4.71**	0.000
List 2 Recall 2 – List 3 Recall 2	0.08	0.32	0.24	0.810
List 3 Recall 2 – List 4 Recall 2	0.98	0.31	3.18*	0.003
List 1 Recall 2 – List 4 Recall 2	2.60	0.34	7.66**	0.000

*$p < 0.01$, **$p < 0.001$
**M = mean sum for each row*

Table 7

Repeated measure t-test results for prediction magnitudes within trials

	M	SEM	*t-test*	*p-value*
List 1 Prediction 1 – List 1 Prediction 2	1.04	0.38	2.74*	0.009
List 2 Prediction 1 – List 2 Prediction 2	1.19	0.51	2.33	0.24
List 3 Prediction 1 – List 3 Prediction 2	1.23	0.46	2.67*	0.010
List 4 Prediction 1 – List 4 Prediction 2	1.94	0.49	4.01**	0.000

*$p < 0.01$, **$p < 0.001$
**M = mean sum for each row*

Table 8

Repeated measure t-test results for first prediction magnitudes between trials

	M	SEM	*t-test*	*p-value*
List 1 Prediction 1 – List 2 Prediction 1	0.89	0.49	1.80	0.078
List 2 Prediction 1 – List 3 Prediction 1	0.35	0.51	0.68	0.501
List 3 Prediction 1 – List 4 Prediction 1	0.42	0.45	0.95	0.348
List 1 Prediction 1 – List 4 Prediction 1	1.65	0.52	3.19*	0.002

*$p < 0.01$, **$p < 0.001$
**M = mean sum for each row*

Table 9

Repeated measure t-test results for second prediction magnitudes between trials

	M	SEM	*t-test*	*p-value*
List 1 Prediction 2 – List 2 Prediction 2	1.04	0.49	2.13	0.038
List 2 Prediction 2 – List 3 Prediction 2	0.39	0.41	0.95	0.347
List 3 Prediction 2 – List 4 Prediction 2	1.14	0.43	2.64	0.011
List 1 Prediction 2 – List 4 Prediction 2	2.56	0.53	4.83**	0.000
List 2 Prediction 2 – List 3 Prediction 2	1.52	0.44	3.46*	0.001
List 1 Prediction 2 – List 3 Prediction 2	1.42	0.56	2.55	0.014

*$p < 0.01$, **$p < 0.001$
**M = mean sum for each row*

Discussion

Hypothesis one posed that that children's overconfidence would decrease within trials in both the experimenter monitored group and the self-assisted monitored group (replicating Lipko-Speed, 2013). Children would give less overconfident second performance predictions for a set of pictures, compared to their first performance predictions for the same set of pictures. Results showed that there was no differences found between groups. The only significant replication of Lipko-Speed's (2013) pattern of decreased overconfidence was found in Trial 4. However, replication was found when children's prediction magnitudes were analyzed rather than overconfidence. Consistent with Lipko-Speed's (2013) findings, children's performance prediction magnitudes decreased within trials. Thus, children provided a lower estimate of the number of pictures they would be able to recall the second time they predicted their performance for a set of pictures.

Hypothesis two stated that children in the self-assisted monitored group would give lower recall predictions between trials for different sets of pictures. By assisting in their own performance monitoring feedback, it was thought that children would be less overconfident. Results showed that children's first predictions did decrease across trials - children predicted they would remember fewer pictures when presented with a new set of pictures, but this was only significant between Trials 1 and 3 and between Trials 1 and 4. This could suggest that children were lowering their predictions on purpose and learning from their past experiences because their predictions were decreasing between some trials. However, they were not decreasing their predictions in all sequential trials. If children were truly using the knowledge that the past predicts the future, their first predictions would have decreased between Trials 2 and

3 and between Trials 3 and 4 as well. This was not found. Instead, children's first predictions were decreasing in an inconsistent manner.

Second predictions also decreased across trials (except between Trials 2 and 3) in that children predicted they would recall less pictures when asked to make a judgment for different sets of pictures. Children were making adjustments in their predictions, yet findings determined that children's second predictions on a trial were not significantly different than their first predictions on the subsequent trial. Specifically, immediately following children's second judgment predictions, their first prediction when presented with a new set of pictures were not significantly different than their second prediction in the previous trial. This provides further support that children were not purposefully decreasing their predictions based on experience.

Even though children's predictions decreased within all and between some trials, their overconfidence scores did not. Because children were recalling fewer pictures every time they were presented with a new set of pictures, overconfidence remained high. Children were still overconfident because both prediction and recall magnitudes were declining together. The implementation of this investigation's self-assisted monitoring task did not aid in decreasing children's overconfident judgment predictions on future tasks.

It was intended for the stars used in the self-assisted monitored group to aid children in making them better predictors of their performance abilities. Physically monitoring their performance by moving a star each time they accurately remembered a picture, it was thought that children would be better able to keep track of their performance and apply their knowledge of their past ability to a new but similar task. Dunlap and Dunlap (1989) used a similar self-monitoring task to increase students' mathematic skills. Children were given a daily checklist where they were to record a plus for each step they did correctly or a minus for each step they did not get right on a subtraction worksheet. Researchers found that when self-monitoring checklists were implemented, the amount of right answers for each subtraction worksheet increased across trials. This study implies that by giving children a physical task to keep track of and monitor their own mistakes and correct steps, they were able to adjust their future performances based on their previous assessments. Monitoring helped children in Dunlap and Dunlap's (1989) study, however in the current study, the self-assisted monitoring implementation did not help children use their past performance ability to predict their future.

A possible reason why Dunlap and Dunlap's (1989) self-monitoring implementation was effective and not the current study's monitoring implementation may be due to the children's age. Even though Dunlap and Dunlap's (1989) participants had learning disabilities, they were still much older than the children in this study. Four- and five-year-old children may be too young to benefit from a self-monitoring aid. The self-assisted monitoring task in this investigation may have been too difficult. The children may not have the cognitive capability yet to understand that using a monitoring device will help them in predicting their future performance. With the memory task used in Lipko-Speed's (2013) study already being cognitively straining, it also may have taken children longer to understand how to play the game because there were more components to it than her original methods. Task difficulty may have influenced the results where if children had more time to understand all of the components of the game, they may have understood what the stars were used for. Conversely, fatigue also may have affected the results in that children became tired after the second trial. This may have caused the increase in overconfidence seen between Trials 3 and 4. Children were more overconfident when making a prediction for a new set of pictures than their first recall prediction in the first trial. Children were getting agitated and tired mid-way through the protocol, which may have led them to think unclearly and not focus on the task at hand. Furthermore, the handling of the stars by the experimenter and the children may have distracted the participants. While playing the game, children were more focused on keeping track of and playing with the stars than they were on recalling the pictures. Using the stars may have also interrupted the children's recall. Instead of children themselves remembering to take a star out of the bowl every time they accurately recalled a picture, many times in the self-assisted monitored group, children had to be prompted to do so. This may have influenced the children's ability to remember as many pictures as they could have because the experimenter interfered with their recall. Intended to be a monitoring aid, the stars seemed to have impacted the children's overconfident metacognitive judgments in the opposite way the current investigation anticipated.

If a different, less cognitively demanding metacognitive memory task were used instead of Lipko-Speed's (2013) picture recall task, the children may have had more available mental resources to better understand what the stars were intended for and would have been able to use them appropriately. Alternatively, if children were trained to use the stars as an aid with a different task and then given the opportunity to complete the picture recall task using the stars, perhaps the stars would have had the intended effect. The cognitive load would be minimized if children were not learning how to use the stars while simultaneously completing the picture recall task. It is also possible that the stars may be a more appropriate aid for a physical task like the ones used in Schneider (1998), rather than a cognitive task like the one used in the current investigation. This would be interesting to investigate in future explorations.

Investigations like the current study in which metacognitive development is researched have important implications for children's learning. Children who are overconfident may experience negative consequences as the result of their overestimations. (Stipek & Mclver, 1989). For example, if children feel confident that they have learned material successfully or mastered a task, they will likely stop studying or practicing. However, if their true mastery of the skill or material does not match their estimation, their performance will likely be poor (Lipko, Dunlosky, Rawson, Swanm & Cook, 2007). Understanding ways to help children be more accurate in their estimations of their own abilities can increase the likelihood that children experience success and mastery of tasks and material in a learning environment.

Reference

Broden, M., Hall, R. V., & Mitts, B. (1971). The effect of self-recording on the classroom behavior of two eighth-grade students. *Journal of Applied Behavior Analysis*, *4,* 191-199.

Clifford, M. M. (1978). The effects of quantitative feedback on children's expectation of success. *British Journal of Educational Psychology*, *48,* 220-226.

Dunlap, L. K., & Dunlap, G. (1989). A self-monitoring package for teaching subtraction with regrouping to students with learning disabilities. *Journal of Applied Behavior Analysis*, *22*, 309-314.

Flavell, J. H., Friedrichs, A. C., & Hoyt, j. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, *1*, 324-340.

Hallahan, D. P., Lloyd, J. W., Kneedler, R. D., & Marshall, K. J. (1982). A comparison of the effects of self-versus teacher-assessment of on-task behavior. *Behavioral Therapy*, *13,* 715-723.

Kazdin, A. E. (1974). Reactive self-monitoring: The effects of response desirability, goal setting, and feedback. *Journal of Consulting and Clinical Psychology*, *42*, 704-716.

Lipko-Speed, A. R. (2013). Can young children be more accurate predictors of their recall performance? *Journal of Experimental Child Psychology*, *114*, 357-365.

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

Lipko, A. R., Dunlosky, J., Rawson, K. A., Swan, K., & Cook, D. (2007). *Using feedback to improve undergraduate and middle-school students' metacomprehansion accuracy.* Kent State University: Unpublished manuscript.

Lyon, L. S. (1983). A behavioral treatment of compulsive lip biting. *Journal of Behavioral, Therapeutic and Experimental Psychiatry*, *14*, 275-276.

Powel, W. D., Bolich, C., & Stewart, M. (1993). Changing predictions based on knowledge of past performance in preschool and young grade school children. *Child Study Journal, 23,* 209-227.

Powel, W. D., Morelli, T., & Nusbaum, N. (1994). Performance and confidence estimates in preschool and young grade-school children. *Child Study Journal*, *24*, 23-47.

Serra, M. J., & Dunlosky, J., (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1258-1266.

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

Shin, H., Bjorklund, D. F., & Beck, D. F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development*, *22*, 197-212.

Stipek, D. J., & Maclver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, *60*, 521-538.

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*, 19983-1989.

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.

Appendix

Pictures used for tasks:

Trial 1:

- House
- Grapes
- Bike
- Cat
- Flowers
- Shirt
- Present
- Dress
- Pizza
- Shark

Trial 2

- Shoes
- Cookies
- Dog
- Gloves
- Keys
- Bear
- Train
- Doll
- Pencil
- Bug

Trial 3

- Plane
- Balloon
- Sunglasses
- Lion
- Leaves
- Heart
- Fish
- Banana
- Fork
- Lamp

Trial 4

- Cup
- Light bulb
- Comb
- Rocks
- Folders

- Coat
- Cherries
- Pants
- Towel
- Computer