Arner, T., Graham, E., Baranski, M., Al Harthy, I. S., & Was, C. A. (2022). Persistent overconfidence in young children: Impact of magnitude and peer modeling. *Educational Research: Theory and Practice, 33*(3), 1-18.

Persistent Overconfidence in Young Children: Impact of Magnitude and Peer Modeling

Tracy Arner[1], Erin Graham[2], Michael Baranski[3], Ibrahim S. Al Harthy[4], & Christopher A. Was[5]

[1]Arizona State University USA
[2]Gustavus Adolphus College USA
[3]California University Pennsylvania USA
[4]Sultan Qaboos University, Muscat, Oman
[5]Kent State University USA

Abstract: *Young children frequently exhibit persistent overconfidence when evaluating their own physical and cognitive abilities. Although task persistence due to overconfidence may be beneficial in some instances, it may also have adverse effects on academic performance. For example, children may reduce cognitive effort because they mistakenly believe that they understand novel tasks. The present work examined two contexts under which young children make more accurate predictions regarding their own ability to recall previously presented items: peer modeling and the number of items to be recalled. Two experiments using the picture recall paradigm demonstrated that both peer modeling and item magnitude may positively impact prediction accuracy and decrease overconfidence in young children. Despite the impact of these interventions on children's recall predictions as a whole, most children were still overconfident in their judgments of learning (JOL). Thus, lending support to our hypothesis that targeted interventions may reduce, but not eliminate, children's overconfidence in their own cognitive abilities.*

Key Words: metacognition, metamemory, judgments of learning (JOL), persistent overconfidence

Metacognition refers to the knowledge of and control over one's own cognition (Flavell, 1979) and is often considered a higher order, self-reflective cognitive process with monitoring and regulatory components (Schneider, 2011). Three facets of metacognition have been identified and studied in the literature: metacognitive knowledge, metacognitive monitoring, and metacognitive control (Dunlosky & Metcalfe, 2008). Metacognitive knowledge refers to one's declarative knowledge about one's own cognition; metacognitive monitoring is one's evaluation of an ongoing cognitive process; and metacognitive control refers to the control and regulation of ongoing cognitive activity, usually based on the result of metacognitive monitoring. Multiple frameworks have been developed to explain how metacognition functions and how it relates to other cognitive functions. Among them is Nelson and Narens' (1990) two-central-dimension framework which stipulates two interactive levels of cognitive processes: the meta-level (a mental model) and the

Correspondence concerning this article should be addressed to Christopher A. Was, E-mail: cwas@kent.edu.

object-level (the dynamic process). In this model, the flow of information runs from the object-level to the meta-level when monitoring and from the meta-level to the object-level when modification is needed based on the monitoring results so that control can be applied. Their framework suggests that people are capable of monitoring and managing thinking with a higher-level representation (Dunlosky & Metcalfe, 2008). Some models of developmental metacognition prefer to situate the construct within the larger framework of self-regulation, which may also include processes related to the regulation of emotion, motivation, social factors, and cognition (see Whitebread et al., 2009). These models, developed in the Vygotskian tradition, highlight the importance of social interactions and adult support on children's metacognitive behaviors. For example, work by Whitebread and colleagues (2007) found that both the frequency and quality of young children's metacognitive behaviors were impacted by the extent of adult direction and whether or not children were working in a group setting. Research in this tradition has been generally optimistic about the metacognitive abilities of young children, arguing that accurate metacognition is possible in children as young as 3 years old (Schneider & Pressley, 1997; Whitebread et al., 2010).

Despite the argument that young children are capable of monitoring their thinking and learning, there is an abundance of literature indicating that they are persistently overconfident when making evaluations of their own cognitive abilities and estimations of their own knowledge (e.g., de Bruin & van Gog, 2012; Flavell et al., 1970; Yussen & Levy, 1975). Although some evidence suggests that this persistent overconfidence may be adaptive because it supports persistence on difficult tasks (Bjorklund & Bering, 2002; Stipek, 1984) which can lead to subsequent learning (Schwebel & Plumert, 1999; Shin et al., 2007), excessive overconfidence can actually decrease learning. This is because recognizing and correcting errors depends, at least in part, on accurately assessing one's performance (Dunlosky & Rawson, 2012; Hartwig & Dunlosky, 2017). One common method of studying children's (and adult's) self-assessment of knowledge and abilities is through the use of judgments of learning (JOLs). For example, in the commonly used picture recall paradigm, a child is asked to study a set of 10 pictures and then make a JOL predicting how many of those pictures she or he will be able to recall on a subsequent test. As noted above, children's JOLs are often miscalibrated in that they are typically much higher than actual recall. Thus, research in this domain has focused on the goal of reducing or correcting judgments to improve calibration and subsequent cognitive performance (see de Bruin & van Gog, 2012 for a review). Interventions which can support increased accuracy of JOLs may be valuable in the classroom as they may improve learning outcomes (de Bruin, & van Gog, 2012; Dunlosky & Thiede, 2013). The goal of the present study is to evaluate two classroom-based interventions designed to reduce young children's overconfident JOLs.

The first intervention is based on work in the areas of social modeling and observational learning. We hypothesized that, when using the picture recall paradigm, watching a peer adjust their JOL to improve accuracy will help children make more accurate JOLs when they complete the same task. The second intervention involves manipulating the number of items to be recalled. We hypothesized that, as the number of items to be recalled increases, children will recognize the increase in task (recall) difficulty and make more proportionally accurate JOLs. Extant literature describes mixed results with previous intervention attempts aimed at improving the accuracy of children's judgments of learning. Interventions with some success include adjusting the timing of JOLs, feedback after a JOL, and memory for prior exposure to content suggest that interventions may improve JOL accuracy. However, even when interventions do improve JOL accuracy, young children typically remain overconfident in their predictions of recall ability (Lipko, Dunlosky, Hartwig et al., 2009; Lipko, Dunlosky, & Merriman, 2009; Shin et al., 2007; Was & Al-Harthy, 2018). Given that those prior interventions have not successfully improved the accuracy of young children's JOLs, the present work introduces two previously unstudied interventions: peer modeling and set magnitude. Relevant literature for each of the prior interventions is described below.

Timing of Judgments of Learning

The timing of the JOL (immediate or delayed) may be a factor in children's accuracy. An immediate JOL occurs when the judgment is elicited concurrently with the presentation of the to-be-learned item. Delayed JOLs are those elicited after some time has passed since the learning event. Researchers found that elementary school children made less accurate JOLs immediately after completing a task than they did following a delay (Schneider et al., 2000). The *delayed-JOL effect* has also been observed with adult learners (Bui et al., 2018) and preschool children (Lipowski et al., 2013). Researchers suggest that immediate JOLs may be less accurate because they rely on maintenance and retrieval of information from primary memory which, in turn, disrupts an evaluation of the content that is stored in long-term memory. Conversely, delayed JOLs allow children to recall previously studied items from long-term memory. Indeed, if they are able to recall information from long-term memory, they provide a higher JOL. If they are unable to recall information from long-term memory, they predict they will not be able to recall it later (Rhodes & Tauber, 2011). However, it should be noted that even in the cases where judgments are made after a delay, young children still exhibited overconfidence (Lipowski et al., 2013).

The Role of Feedback in Judgments of Learning

Feedback may support learners in adjusting JOLs and predictions of performance. Middle school children adjusted their performance evaluations when provided with explicit standards and idea units (Lipko, Dunlosky, Hartwig et al., 2009). Research investigating the efficacy of explicit feedback with younger children is less consistent. To wit, Lipowski et al., (2013) provide some evidence that preschool students can use explicit feedback to better calibrate their JOLs. Alternatively, work by Lipko, Dunlosky, & Merriman (2009), using the picture recall paradigm, suggests that young children do not make more accurate judgments after explicit feedback. Lipko, Dunlosky, & Merriman (2009) tasked children with studying a series of ten pictures for a brief period of time, then making a prediction about the number of items they would recall after the pictures are hidden, and subsequently recalling as many pictures as possible. In one experiment, (Lipko-Speed, 2013) researchers explicitly told children how many items they had correctly recalled after each trial in an attempt to increase prediction accuracy on the next trial. Results indicated that explicit feedback on their performance was not sufficient to reduce children's overconfidence in their recall ability. It should be noted that, as in the case of delayed JOLs, even when explicit feedback does have a measurable impact on the accuracy of the judgment, many young children remain overconfident (Lipko, Dunlosky, & Merriman, 2009; Lipowski et al., 2013).

A few trends emerge from the studies reviewed. First, although several interventions have been successful in getting children to adjust their JOLs, they are still largely overconfident relative to actual performance. Second, younger children are more likely to generate and maintain overconfident judgments of performance across interventions than older children (van Loon et al., 2017). Third, interventions designed to improve young children's JOLs are more likely to be successful when they provide sufficient external cues, do not overtax cognitive resources, and allow sufficient time for feedback integration (Kloo & Perner, 2008).

The present work incorporates this information into the design and evaluation of two interventions meant to diminish young children's overconfident JOLs. Both interventions were designed to capitalize on external cues. The first incorporates attempts to provide sufficient external cues via observational learning through modeling and the second manipulates perceived task difficulty by varying the quantity of recall items.

Social Modeling and Observational Learning with Developmental Judgments of Learning

The observational learning paradigm has been foundational in social psychology and social cognition (Bandura et al., 1961; Schunk, 1989). One study investigating social modeling with preschool children demonstrated that 3-year-olds may learn category sorting rules implicitly from watching an adult model (Williamson et al., 2010). Similarly, Wang and colleagues (2015) demonstrated that 4-year-old children implicitly learned sorting rules derived from invisible properties of objects (e.g., weight) via modeling. While the role of social factors in developmental metacognition is a popular topic for observational research (see Whitebread et. al., 2010), peer modeling has been sparsely applied in experimental investigations of JOLs in young children. Studies that have incorporated elements of social modeling have found some benefit to children's ability to accurately predict their own learning. For example, one study investigated the impact of providing peer norms (e.g., "your friends were able to recall this many") with preschool, third grade and college-age participants using the picture recall paradigm (Yussen & Levy, 1975). Stimuli sets increased or decreased from one to ten pictures (i.e., trial 1 is one picture, trial 2 is two pictures, so on, up to ten pictures or the reverse). Results suggest that telling participants whether or not their peers could recall all of the items in a trial benefited calibration of JOLs for third grade and college students. Preschool students remained overconfident and unaware of the relation between peers' and their own ability. Yussen and Levy (1975) suggest that preschoolers do not understand that peer performance has implications for their own performance. However, third grade students do understand this implication. We speculate that the development of this understanding is likely influenced by spending more time with peers in settings in which performance in such activities is public and evaluated, such as classrooms. More recently, researchers used direct modeling to investigate whether or not preschool children may adjust their recall response after viewing a model do so. Participants observed an adult who was shown two toys and told that one of the toys was going to be hidden in an opaque box. The toy was then moved either to a location where the model could see the toy being hidden or to a location where the model could not see the toy being hidden. Once hidden, children in the experimental condition observed the model respond correctly to where the toy was located with an explanation for their response (e.g., "Yes, because I saw you put the toy in the box"). The children completed the same task after observing the model. Results indicated that children who observed the model provide the correct answer with an explanation were more likely to accurately assess and correctly explain their own knowledge (Lipko-Speed et al., 2018).

Task Difficulty and Judgments of Learning

Growing evidence suggests that perceived task difficulty may affect the accuracy of children's judgments of learning (Koriat & Shitzer-Reichert, 2002; Was & Al-Harthy, 2018). Recent research found that fourth-grade students successfully used task difficulty to calibrate their judgments of learning such that students consistently decreased predictions for more difficult items whereas easier items were consistently given a higher judgment of learning. These results suggest that older children are capable of calibrating judgments of learning when given tasks with sufficiently high perceived difficulty. However, the same study found that second grade students were still consistently overconfident when generating judgments of learning for difficult items (Koriat & Shitzer-Reichert, 2002). Alternatively, Destan et al., (2014) found a significant difference in JOLs between easy and difficult items when testing 5- to 7-year-olds with all children providing higher JOLs to easy items compared with difficult items. The authors concluded that even 5-year-olds were able to successfully monitor the differential difficulty of the task.

Perceived task difficulty was also investigated using the picture recall paradigm that included familiar and unfamiliar items with fourth- and sixth-grade students. The researchers hypothesized that students would be more confident in their ability to recall familiar (less difficult) items and less confident in their ability to recall unfamiliar (more difficult) items (Was & Al-Harthy, 2015). The results of this study provided some evidence that perceived task difficulty did positively affect (i.e., increased accuracy) children's JOLs. The effect was much smaller than expected and children remained overconfident in their predictions regardless of stimuli familiarity. Interestingly, one study demonstrated that young children are capable of incorporating task difficulty (instantiated as item familiarity) into their JOLs, but that this calibration does not result in long term changes in judgment accuracy (Lipko-Speed, 2013). Preschool students were asked to make two judgments of learning for each set of items in the picture recall paradigm; a standard, initial prediction followed by a second prediction immediately following their attempted recall. Results indicated that the children were less overconfident (i.e., generated a lower prediction) when making their second prediction, suggesting that the children integrated perceived task difficulty with predictions of their performance when making the second JOL. However, as soon as the next trial began and the recall materials changed, the children returned to making overconfident predictions. The researcher concluded that children could integrate their own prior difficulties with a task when generating judgments of learning, but this only leads to a small, temporary decrease in their usual overconfidence (Lipko-Speed, 2013).

EXPERIMENT 1

Experiment 1 of the present study seeks to build on the work of Lipko-Speed et al., (2018) and Lipko, Dunlosky & Merriman (2009) by combining the picture recall paradigm with social modeling to improve the accuracy of young children's judgments of learning. Recall that children in Lipko, Dunlosky & Merriman (2009) were given corrective feedback after predicting their recall ability in the picture recall paradigm. However, they were not shown a social model actively incorporating the corrective feedback into subsequent predictions. In contrast, children in Lipko-Speed et al., (2018) were exposed to a social model making correct decisions based on explanations, but they were not asked to make JOLs (such as in the picture recall paradigm) nor did they see the model incorporating corrective feedback. Additionally, the present work includes a peer model (i.e., 5-year-old female student) rather than an adult model. Experiment 1 of the present study combined these ideas and investigated whether witnessing a peer model incorporating corrective feedback when generating judgments of learning resulted in more accurate predictions from young children on completing the same task. We hypothesized that preschool children who observe a peer model incorporate corrective feedback would adjust their own judgments of learning and demonstrate reduced overconfidence.

Method

Participants

A conservative a priori power analysis with an effect size of Cohen's $f = .53$ (drawn from the extant literature), alpha = .05, beta = .95 suggested a total sample of thirty-two participants. Sixty students (M_{age} = 5.5 years, SD = .26, R = 5.11 - 5.90, 60% female) in first grade from a public elementary school in Muscat, Oman were randomly selected from different classrooms to participate in this study. The elementary school cooperates with the College of Education, Sultan Qaboos University in many matters such as teacher education field experiences, teacher training, and educational research. The public school system in Oman does not collect data regarding race or ethnicity, as the government does not reinforce identification of this type of diversity. All ethical standards of the American Psychological Association were followed in the treatment of participants and collection of the data. Participants were randomly assigned to either a correction or no correction condition.

Materials and Procedure

Students in both conditions were invited to play a memory game in which they were asked to identify and then recall 10 familiar items (Figure 1) from small picture cards. The picture cards were 3" x 3" cards with full color drawings of the familiar items.

Figure 1
Picture cards of common items participants were asked to remember

Following assent, students were shown a video of a peer demonstrating the game. The model for all participants in both conditions was a 5-year-old, female student. Students in the no-correction condition viewed the model complete two trials. In the first trial, the experimenter had the model name the pictures on ten cards and asked her to predict how many items she would be able to recall after the experimenter covered the cards. The model predicted that she would recall

all 10 of the items after the cards were covered, and subsequently recalled only four items. The experimenter in the video then conducted a second trial with the model. On the second attempt, the model had the same overconfident prediction (that she would recall 10 items) and then again recalled only four. The correction condition video was the same except that in the second trial, the model adjusted her prediction to reflect her recall from the first trial (4 items) and again recalled four items. After viewing the short video demonstration of the memory game, students were again asked if they would like to play[1].

Participants in both conditions were shown picture cards of familiar items (Figure 1) and asked to identify each item by name. If students were unable to name an item, it was replaced with one they could identify (all children were able to identify all items on the cards without errors). After participants were shown the cards and correctly identified 10 items, they were asked to predict how many items they would be able to recall after the cards were covered. Participants were then given 20 seconds to study the pictures. After the cards were covered, participants were asked to recall as many items as they could. The same procedure was repeated with a novel set of 10 picture cards for Trial 2. After participants completed both trials, they were thanked and given a small gift.

RESULTS

Our analytic plan consisted of a series of factorial ANOVAs. To begin, we conducted a 2 (between subjects: no-correction vs. correction) x 2 (within subjects: Trial 1 vs. Trial 2) factorial ANOVA on prediction overconfidence (calculated as predicted number of items minus actual number of items recalled, so that a value of zero would represent perfect accuracy). Figure 2 represents prediction overconfidence by condition and trial. Inspection of the figure suggests that participants in the correction condition were less overconfident than those in the no-correction condition. The results of the factorial ANOVA show that the effect of condition was significant, $F(1, 58) = 33.89$, $Mse = 5.63$, $p < .001$, $\eta^2_{partial} = .37$. The within subjects contrasts were not significant, $F < 1$, nor was the condition by trial interaction, $F(1, 58) = 3.22$, $Mse = 3.74$, $p < .078$, $\eta^2_{partial} = .05$. Thus, the results of this analysis indicate that the participants in the correction condition were significantly less overconfident in their predictions.

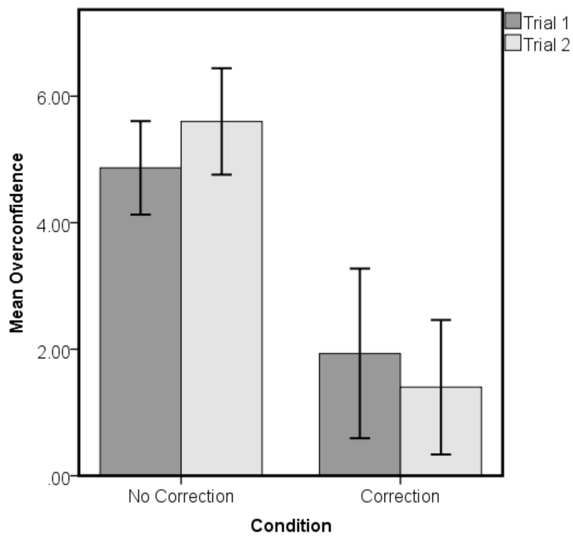
To further understand these results, we ran two more factorial ANOVAs: One using prediction data and another using recall data. Table 1 presents the means and standard deviations of the predicted number of items to be recalled and the number of items actually recalled by condition and trial. The descriptive statistics suggest that participants in the correction condition had lower predictions, but did not differ in the number of items recalled than those in the no-correction condition. The 2 (between subjects: no-correction vs. correction) x 2 (within subjects: Trial 1 vs. Trial 2) factorial ANOVA using predictions demonstrated a significant effect of condition, $F(1, 58) = 85.49$, $Mse = 3.12$, $p < .001$, $\eta^2_{partial} = .60$. There was also an effect of trial, $F(1, 58) = 10.24$, $Mse = 2.12$, $p = .002$, $\eta^2_{partial} = .15$, and a marginal interaction, $F(1, 58) = 3.78$, $Mse = 3.12$, $p = .057$, $\eta^2_{partial} = .06$. These results indicated that participants in the correction condition made significantly lower predictions than those in the no-correction condition.

Last, we conducted a similar 2 x 2 factorial ANOVA using recall data. The results did not indicate an effect of condition, $F(1, 58) = 3.66$, $Mse = 1.73$, $p = .060$, $\eta^2_{partial} = .06$. There was a significant effect of trial, $F(1, 58) = 14.54$, $Mse = 1.86$, $p < .001$, $\eta^2_{partial} = .20$, but no interaction,

[1] Videos available by request from cwas@kent.edu

F < 1. These results suggest that participants in the two conditions did not differ in their actual recall of the picture-card items.

Figure 2
Means of Overconfidence by Condition and Trial



Note. Error bars represent 95% confidence interval. Overconfidence was calculated as Prediction Minus Recall.

Table 1
Means and Standard Deviations for Prediction and Prediction Proportion by Trial and Condition

	Prediction 1		Recall 1		Prediction 2		Recall 2	
Condition	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
No-Correction	9.73	.91	4.87	1.85	9.40	1.43	3.80	1.67
Correction	6.03	2.83	4.10	1.49	4.67	2.41	3.27	1.28

Discussion

The results from Experiment 1 indicate that participants in the correction condition made lower and thus, more accurate predictions of the number of items they would be able to recall during two trials of the picture recall paradigm compared to participants in the no-correction condition. In support of our hypothesis, this suggests that participants in the correction condition were able to imitate the behavior of the model and adjust their predictions of their own recall ability. This aligns with other studies that incorporated elements of social modeling to increase children's ability to accurately predict their own learning. Further, there was no difference in recall between conditions. Therefore, we are confident that watching the model adjust their prediction for the second trial, was informative and participants in the correction condition imitated the model when making their predictions in the first (and second) trial.

One implication of the findings from the experiment is they suggest that children are able to modify their predictions about their ability to memorize and recall information with the appropriate cues. In this case, the cue was an age-appropriate peer, engaged in the same task. When young children are challenged to understand their own range of abilities, providing a peer model might enhance a child's ability to assess their own possible range of performance. We also recognize that the experiment has certain limitations, for example the peer model was a child on a video and not a classmate, yet the findings are promising.

The results of Experiment 1 suggest that young children can learn to improve their memory prediction overconfidence via modeling. Experiment 2 investigated conditions under which students may be able to calibrate their predictions independently, without cues from an external source. One condition that may lead to better calibration is perceived task difficulty.

EXPERIMENT 2

Experiment 2 of the present study sought to build on prior work investigating perceived task difficulty (Koriat & Shitzer-Reichert, 2002; Lipko-Speed, 2013; Was & Al-Harthy, 2015) by exploring whether a new instantiation of task difficulty will be sufficient to counteract children's persistent overconfidence when generating JOLs. Specifically, Experiment 2 examines the extent to which the quantity of to-be-remembered items (5, 10, or 15) presented during the picture recall paradigm influences overconfidence and accuracy of children's judgments of learning. We hypothesized that children would perceive the larger item sets as being more difficult to remember and decrease their predicted proportion of recall, thus reducing their overconfidence.

In the extant literature regarding young children's ability to predict their performance on a memory task – specifically the literature using the picture recall paradigm – the number of items to be studied and recalled is typically 10. Studies using 10 items in this paradigm demonstrate that young children are overconfident in their ability to recall the pictures presented (i.e., on average predicting about 8) when their actual recall is usually quite poor (i.e., on average about 4 items). It is therefore unclear if the number of items presented is a source of the overestimation of the number of items the children believe they can recall. To understand the relationship between the number of items presented and children's ability to calibrate, they must be presented with the opportunity to differentiate task difficulty. We hypothesized that young children could differentiate task difficulty when presented with a variety of difficulty levels and thus we propose the perceived difficulty hypothesis. The perceived difficulty hypothesis predicts that as the number of items presented for recall increases, so does perceived difficulty. In turn, the proportion of items predicted to be recalled by the participants will decrease. To test this hypothesis, we had children complete three trials of the picture recall paradigm. The trials consisted of five, ten, and fifteen items with set-size order counterbalanced across participants.

METHOD

PARTICIPANTS

An a priori power analysis with an effect size of Cohen's f = .25, alpha = .05, beta = .80 suggests a total sample of twenty-eight participants. Twenty-nine (M_{age} = 5.3 years, SD = .48, R = 4.11 - 5.90, 69% female) first graders from the same public elementary school who did not participate in Experiment 1 were randomly assigned to one of six conditions where the picture card quantity presentation order was varied to manipulate perceived task complexity. Table 2 presents the number of participants in each counterbalanced condition. Due to random assignment error, we did not have an even number of participants in each condition (see Table 2). As in Experiment 1, participants were presented with picture cards (Figure 1) and asked to identify each of the items on the cards. Participants were unaware of the number of cards in each set until they were asked to identify the items for each trial. When the required number of cards were successfully named, participants were given 10, 20, and 30 seconds to study them, respectively. Participants were again asked to predict how many items they would recall after the cards were covered. After covering the cards, participants were asked to recall as many items as possible. The same procedure was followed with the next quantity of cards, dependent upon condition, until all three quantities were completed. After participants completed all three prediction/recall trials, they were thanked and given a small gift.

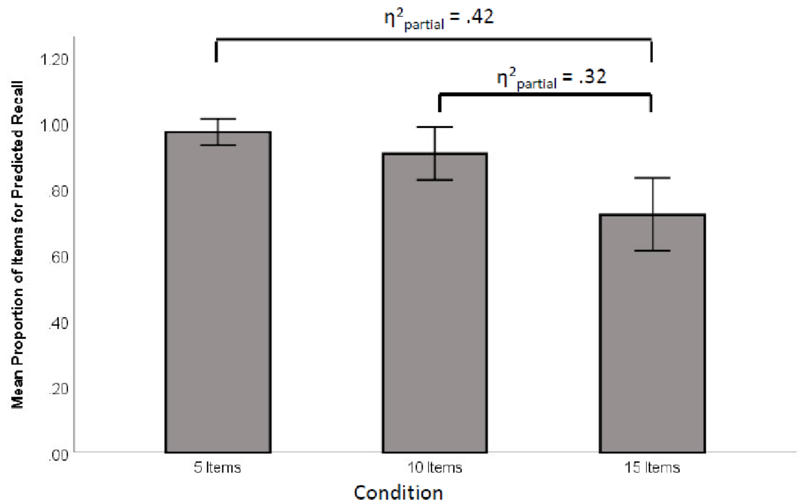
Table 2
Number of Participants and Presentation Order by Condition

Condition	Participants (*N*=29)	Presentation Order
1	6	5, 10, 15
2	4	5, 15, 10
3	6	10, 5, 15
4	5	15, 5, 10
5	5	10, 15, 5
6	3	15, 10, 5

Results

Recall that the perceived difficulty hypothesis states that as the number of items presented increases, the proportion of items predicted to be recalled will decrease. Table 3 presents the means and standard deviations for the number and proportion of items predicted by the number of items presented. To test the hypothesis, we conducted a repeated measures ANOVA with simple contrast comparing the trials with 15 items to the two other trials. Figure 3 depicts the mean proportion of predicted items to be recalled by number of items. The RM ANOVA indicates that on average, participants did predict a significantly lower proportion of items would be remembered as the number of items presented increased, $F(2, 56) = 14.31$, $Mse = .03$, $p < .001$, $\eta^2 = .34$. Within-subject contrasts indicate that the proportion of items predicted to be recalled was smaller in the 15-item trial compared to both the 5-item trial, $F(1, 28) = 20.29$, $Mse = .09$, $p < .001$, $\eta^2_{partial} = .42$, and the 10-item trial, $F(1, 28) = 13.30$, $Mse = .08$, $p = .001$, $\eta^2_{partial} = .32$.

Figure 3
Means of Proportion of Items for Predicted Recall



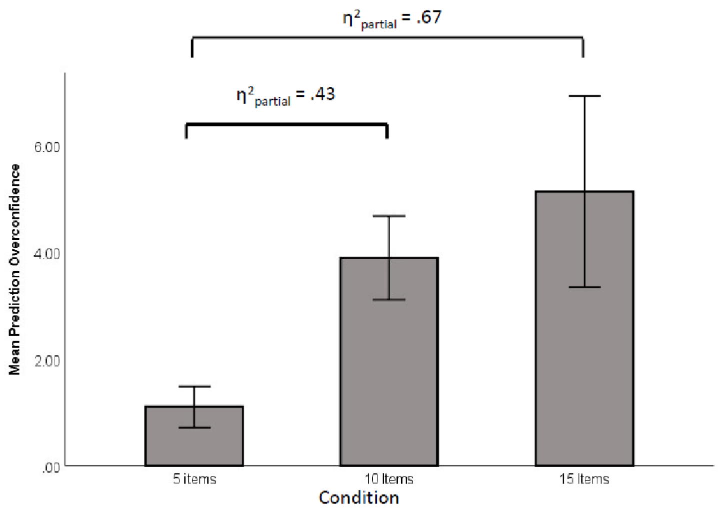
Note. Error bars represent 95% confidence intervals.

As in Experiment 1, we also analyzed prediction overconfidence data by condition. Prediction overconfidence was again calculated as the number of predicted items minus the number of recalled items. Therefore, larger numbers represent greater overconfidence. Figure 4 depicts mean overconfidence by trial type. A RM ANOVA with simple contrast comparing the 5-item trial to the two other trials was significant, $F(2, 56) = 15.89$, $Mse = .78$, $p < .001$, $\eta^2_{partial} = .36$. The within-subjects contrast indicates that when compared to the 5-item trial, overconfidence was significantly worse in the 10-item trial, $F(1, 28) = 53.34$, $Mse = 4.24$, $p < .001$, $\eta^2_{partial} = .67$, and the 15-item trial, $F(1, 28) = 20.69$, $Mse = 22.80$, $p = .001$, $\eta^2_{partial} = .43$.

Table 3
Means and Standard Deviations for Prediction and Prediction Proportion by Condition

	Prediction		Proportion	
Condition	*Mean*	*SD*	*Mean*	*SD*
5 Items	4.86	.52	.97	.10
10 Items	9.07	2.12	.91	.21
15 Items	10.83	4.35	.72	.29

Figure 4
Means of Prediction Overconfidence by Condition



Note. Error bars represent 95% confidence intervals.

Discussion

Although participants in Experiment 2 did decrease the proportion of items they predicted to recall according to set size - supporting our perceived difficulty hypothesis - they were still significantly overconfident in their predictions for larger sets than smaller sets, demonstrating persistent overconfidence. Therefore, we suggest that the perceived difficulty of the task did influence the participants' predictions, yet the effect was not large enough to completely eliminate overconfidence. Put differently, the results of this experiment suggest that young children understand that attempting to recall items from larger sets is more difficult than attempting to recall items from the smaller set, but they remain overconfident in their recall ability. As seen by the mean prediction overconfidence scores, participants were still substantially overconfident after they adjusted predictions for larger set sizes. This may, in part, be an artifact of the data due to the ease of recalling five items (the mean proportion of items recalled for the 5-item condition was M = .75) as compared to recalling 15 items (mean proportion recalled M = .38), although this still represents a good deal of overconfidence. Thus, the results of Experiment 2 suggest that young children are able to recognize the difficulty in attempting to remember and recall more items as compared to fewer items, yet they are still overconfident in their predictions in the more difficult condition.

General Discussion

Decades of research suggests that young children are generally overconfident in their own cognitive capabilities, with both positive and negative consequences (e.g., persistence at difficult tasks; failure to regulate study in line with actual ability). The current research used a common, picture recall paradigm to investigate if young children reduce their overconfidence and better calibrate recall predictions after observing a peer model adjust predictions in line with actual recall (Experiment 1) or when considering the difficulty of the recall task (Experiment 2). Results from both experiments suggest that children's prediction accuracy can be improved; however, decreases to predictions were minimal and children remained overconfident.

Peer Information, Modeling, and Feedback

Results of Experiment 1 indicated that watching a video of a peer integrating corrective feedback and adjusting subsequent recall predictions helped children adjust their own recall predictions on the first and second trials. These findings align with and extend prior research regarding peer comparison, modeling, feedback, and prediction accuracy (Yussen & Levy, 1975; Lipko, Dunlosky, & Merriman, 2009; Lipko-Speed et al., 2018). The first graders in the current research developmentally fit between Yussen and Levy's (1975) preschoolers and third graders. Their predictions were more similar to the third graders, who calibrated based on information regarding how much their peers could recall than to preschoolers, who did not calibrate predictions based on similar information.

The current results and those of Yussen and Levy (1975) contrast with Lipko et al. (2009b) regarding the use of recall performance information in calibrating subsequent predictions. Yussen and Levy (1975) found that when participants were given information regarding how many items their peers could recall on particular sets of pictures, they were able to use that information to calibrate their own recall predictions. Similarly, in Experiment 1, participants who viewed a peer model make an overconfident prediction, fail to recall the predicted amount, and then adjust the subsequent prediction in line with prior recall performance were able to adjust their own recall prediction. In contrast to these results, Lipko et al., (2009b) found that while participants could accurately report their previous recall performance, they still failed to integrate this information when making subsequent recall predictions, thus remaining overconfident. Combining the results of these two prior studies with Experiment 1 suggests that young children may need modeling or explicit instruction regarding integration of prior performance with subsequent tasks in order to calibrate their predictions to improve accuracy.

The results of Experiment 1 did cohere with the findings of Lipko-Speed, Buchart, and Merriman (2018) in that more accurate judgments were made after watching a model. Whether this response comes from participants calibrating their own predictions to match their performance or merely imitating the response they observed is unclear. In summary, across prior (Yussen & Levy, 1975; Lipko-Speed et al., 2018) and current research, incorporating task performance feedback, whether given directly regarding a peer's performance or from observing an adult or peer model, improved recall prediction accuracy for first grade and older children. Considering the current results, along with Lipko, Dunlosky, and Merriman (2009), young children may have to be given guidance or directly observe how to use performance feedback to calibrate predictions.

Item Difficulty

Results from Experiment 2 suggest that children can calibrate their immediate recall predictions when confronted with sufficiently difficult tasks. To wit, children in Experiment 2 expected to recall a smaller proportion of 15 pictures than 5 pictures as reflected in their lower recall predictions on larger item sets. This finding aligns with previous research (Destan et al., 2014; Was & Al-Harthy, 2015, 2018) indicating that young children can appropriately calibrate recall predictions based on item or task difficulty. However, even when adjustments do occur, children are still largely overconfident in their predicted recall ability. Current results suggest that young children may be somewhat sensitive to the memory load imposed by the quantity of items to be recalled (5, 10, or 15) when making predictions about their recall ability, but this sensitivity is not enough to create meaningful differences in their persistent overconfidence. This result also aligns with previous research (Koriat & Shitzer-Reichert, 2002; Lipko-Speed, 2013) that children may not completely or consistently calibrate their predictions in line with item or task difficulty. In summary, young children are capable of calibrating recall predictions when confronted with varying item difficulty, whether due to item familiarity or load on memory, yet remain overconfident, nonetheless.

LIMITATIONS

We recognize that the present work is not without limitations. First, though we did achieve power to support our results, the samples were relatively small. Thus, we were unable to complete any post hoc analyses such as the influence of order in Experiment 2. In addition, this work would benefit from replication with larger, demographically diverse samples. Second, the model used in this work was presented via a pre-recorded video. The benefit of vicarious learning may be enhanced when using a classmate to directly model calibration of JOLs. Finally, the students in this work were given feedback on the number of items they recalled correctly; however, students were not provided with an evaluation of the number of cards recalled compared to the number of cards they predicted.

FUTURE DIRECTIONS AND IMPLICATIONS FOR EDUCATION

The results of the current research suggest young children's prediction accuracy is malleable yet they remain overconfidently inaccurate. Relevant aspects of theory and application remain unknown, which opens up several promising lines of research. Regarding the effects of peer modeling on recall prediction calibration examined in Experiment 1, the current research used a two-trial peer modeling calibration with a two-trial prediction/recall of pictures with young children. Avenues for future research include investigating durability of the benefit, effectiveness of model type, developmental trajectory and classroom implementation. These and prior studies do not contain trials after a delay, thus future research should evaluate the duration of effects as well as the number of treatments needed to obtain long-lasting improvements in prediction calibration. Additionally, future research is needed to determine methods to maximize the benefit of peer models. Specifically, investigating differences between direct modeling by a classmate and an unknown model presented electronically may help determine whether the cost of implementation is justified. Furthermore, this work evaluated peer modeling with young children, and it is unknown whether or not older children (i.e., 3rd to 5th grade) would experience the same benefit or whether developmental gains would surpass those gained from a peer-modeling intervention. Additional work to determine effective use of peer modeling calibration in an actual learning environment may promote an accurate understanding of learning, improved prediction of performance thus achieving far-reaching benefit to education as children begin to understand and adapt their own learning behaviors.

Experiment 2 of the current research used three set sizes of pictures to be recalled (5, 10, and 15) to measure the effect of task difficulty on prediction calibration in young children. Future research should contrast very difficult to very easy task predictions (e.g., distinctly large and distinctly small set sizes) to examine the effect on prediction calibration. Additionally, the compound effects of multiple manipulations on prediction overconfidence remain unknown. One avenue for future research is to investigate multiple trials with each of the set sizes to evaluate calibration based on self-reflection for each trial. Recall that students improved the accuracy of their recall prediction when the same set of items was repeated, but this may be due to the practice effect (Lipko-Speed, 2013). Students were also able to adjust predicted recall and confidence in predictions depending on the familiarity of items (Was & Al-Harthy, 2015).

The current experiments add to the literature suggesting that corrective peer modeling and, to a lesser degree, task complexity (instantiated as set size), may positively influence recall predictions. However, further investigation is needed to examine the compound impact of item novelty, multiple exposures to corrective peer modeling, set size, age, and repetition on the calibration of children's recall prediction. Finally, regarding implications for education, a takeaway from the current and previous research is that young learners have some awareness of their own abilities but can benefit from assistance – adult or peer – in modeling accurate judgment of abilities.

Conclusion

Young children exhibit persistent overconfidence in their own cognitive abilities which may be detrimental to learning. Students who believe that they know more than they do may engage in ineffective study practices or fail to study at all (Dunlosky & Rawson, 2012). Classroom interventions that improve the accuracy of children's JOLs may benefit their short- and long-term learning outcomes. The present work investigated two interventions that may fill this need. First, students observed a peer model use prior recall to correct their judgment of learning on the next recall task. Students were then able to integrate the model's performance into the generation of their own JOL. Students in the correction condition, though still overconfident, did decrease their predicted recall on the first and second trials in the picture recall paradigm. This result is promising, and further investigation is warranted as it is a low-cost, easily replicable intervention. Second, results of Experiment 2 suggest young children may be able to calibrate their judgments of learning independently when confronted with a task they perceive to be sufficiently difficult (i.e., large set size). Students in both experiments adjusted their predictions of recall ability following classroom interventions. Improved accuracy in generating judgments of learning may support future learning when students can accurately identify what material they have learned and what material still needs to be studied.

References

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology, 63*(3), 575. doi:10.1037/h0045925

Bjorklund, D. F., & Bering, J. M. (2002). The evolved child: Applying evolutionary developmental psychology to modern schooling. *Learning and Individual Differences*, *12*(4), 347-373. doi:10.1016/S1041-6080(02)00047-X

Bui, Y., Pyc, M. A., & Bailey, H. (2018). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: the "Displaced-JOL effect". *Memory,26*(6), 771-783. doi: 10.1080/09658211.2017.1406523

de Bruin, A. B., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction, 22*(4), 245-252. doi:10.1016/j.learninstruc.2012.01.003

Destan, N., Hembacher, E., Ghetti, S., & Roebers, C. M. (2014). Early metacognitive abilities: The interplay of monitoring and control processes in 5-to 7-year-old children. *Journal of Experimental Child Psychology*, *126*, 213-228. doi:10.1016/j.jecp.2014.04.001

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271-280. doi:10.1016/j.learninstruc.2011.08.003

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, 24, 58-61. doi:10.1016/j.learninstruc.2012.05.002

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906. doi:10.1037/0003-066X.34.10.906

Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology, 1*(4), 324-340. doi:10.1016/0010-0285(70)90019-8

Hartwig, M. K., & Dunlosky, J. (2017). Category learning judgments in the classroom: Can students judge how well they know course topics?. *Contemporary Educational Psychology*, *49*, 80-90. doi: 10.1016/j.cedpsych.2016.12.002

Kloo, D., & Perner, J. (2008). Training theory of mind and executive control: A tool for improving school achievement?. *Mind, Brain, and Education, 2*(3), 122-127. doi:10.1111/j.1751-228X.2008.00042.x

Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy. In *Metacognition* (pp. 1-17). Springer, Boston, MA. doi:10.1007/978-1-4615-1099-4_1

Lipko, A. R., Dunlosky, J., Hartwig, M. K., Rawson, K. A., Swan, K., & Cook, D. (2009). Using standards to improve middle school students' accuracy at evaluating the quality of their recall. *Journal of Experimental Psychology: Applied, 15*(4), 307. doi:10.1037/a0017599

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*(2), 152-166. doi: 10.1016/j.jecp.2008.10.002

Lipko-Speed, A. R. (2013). Can young children be more accurate predictors of their recall performance?. *Journal of Experimental Child Psychology*, *114*(2), 357-363. doi: 10.1016/j.jecp.2012.09.012

Lipko-Speed, A. R., Buchert, S., & Merriman, W. E. (2018). Observing an adult model can cause immediate improvement in preschoolers' knowledge judgments. *Cognitive Development, 48*, 225-234. doi: 10.1016/j.cogdev.2018.09.003

Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Developmental Psychology, 49*(8), 1505. doi:10.1037/a0030614

Nelson, T.O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-173). San Diego: Academic Press.

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin, 137*(1), 131. doi:10.1037/a0021705

Schneider, W. (2011). Memory development in childhood. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (p. 347–376). Wiley-Blackwell.

Schneider, W., & Pressley, M. (1997). Memory development between 2 and 20. Hillsdale, NJ: Erlbaum.

Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task. *Cognitive Development, 15*(2), 115-134. doi: 10.1016/S0885-2014(00)00024 doi:10.1016/S0885-2014(00)00024-1

Schunk, D. H. (1989). Social cognitive theory and self-regulated learning. In *Self-regulated Learning and Academic Achievement* (pp. 83-110). Springer, New York, NY. doi:10.1007/978-1-4612-3618-4_4

Schwebel, D. C., & Plumert, J. M. (1999). Longitudinal and concurrent relations among temperament, ability estimation, and injury proneness. *Child Development*, *70*(3), 700-712. doi:10.1111/1467-8624.00050

Shin, H., Bjorklund, D. F., & Beck, E. F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development, 22*(2), 197-212. doi:10.1016/j.cogdev.2006.10.001

Stipek, D. J. (1984). Young children's performance expectations: Logical analysis or wishful thinking. *Advances in Motivation and Achievement*, *3*, 33-56.

van Loon, M., de Bruin, A., Leppink, J., & Roebers, C. (2017). Why are children overconfident? Developmental differences in the implementation of accessibility cues when judging concept learning. *Journal of Experimental Child Psychology, 158,* 77-94. doi:10.1016/j.jecp.2017.01.008

Wang, Z., Meltzoff, A. N., & Williamson, R. A. (2015). Social learning promotes understanding of the physical world: Preschool children's imitation of weight sorting. *Journal of Experimental Child Psychology*, *136*, 82-91. doi:10.1016/j.jecp.2015.02.010

Was, C. A., & Al-Harthy, I. (2015). Developmental differences in overconfidence: When do children understand that attempting to recall predicts memory performance?. *The Researcher*, *27*(1), 1-5.

Was, C. A., & Al-Harthy, I. S. (2018). Persistence of overconfidence in young children: Factors that lead to more accurate predictions of memory performance. *European Journal of Developmental Psychology, 15*(2), 156-171. doi:10.1080/17405629.2016.1264936

Whitebread D., Almeqdad Q., Bryce D., Demetriou D., Grau V., Sangster C. (2010) Metacognition in Young Children: Current Methodological and Theoretical Developments. In: Efklides A., Misailidi P. (eds) *Trends and prospects in metacognition research*. Springer, Boston, MA. doi:10.1007/978-1-4419-6546-2_11

Whitebread, D., Bingham, S., Grau, V., Pasternak, D. P., & Sangster, C. (2007). Development of metacognition and self-regulated learning in young children: Role of collaborative and peer-assisted learning. *Journal of Cognitive Education and Psychology*, *6*(3), 433-455. doi:10.1891/194589507787382043

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., Almeqdad, Q., & Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning*, *4*(1), 63-85. doi:10.1007/s11409-008-9033-1

Williamson, R. A., Jaswal, V. K., & Meltzoff, A. N. (2010). Learning the rules: Observation and imitation of a sorting strategy by 36-month-old children. *Developmental Psychology*, *46*(1), 57.

Yussen, S. R., & Levy Jr, V. M. (1975). Developmental changes in predicting one's own span of short-term memory. *Journal of Experimental Child Psychology, 19*(3), 502-508. doi:10.1016/0022-0965(75)90079-X