Metacognition Learning (2013) 8:19–46
DOI 10.1007/s11409-013-9094-7

Metacognition and control of study choice in children

Janet Metcalfe · Bridgid Finn

Received: 1 May 2012 / Accepted: 28 January 2013 /
Published online: 12 February 2013

Abstract Middle childhood may be crucial for the development of metacognitive monitoring and study control processes. The first three experiments, using different materials, showed that Grade 3 and Grade 5 children exhibited excellent metacognitive resolution when asked to make delayed judgments of learning (JOLs, using an analogue scale) or binary judgments of knowing (JOKs, 'know' or 'don't know') without the target being present. (The delayed method used here also results in excellent metacognitive resolution in adults). In three subsequent experiments after making JOLs the children were asked to choose which items they would like to restudy to optimize learning. We then either honored or dishonored the children's restudy choices, and tested their memory performance. In Experiment 4, honoring the children's choices made no difference to final recall performance. Experiments 5 and 6 showed that when the computer, rather than the children, chose the items for restudy based on theoretical constraints proposed by the Region of Proximal Learning model of study time allocation, the children's recall performance improved. In all three experiments, Grade 3 children's choices were random. Whereas the Grade 5 children showed some indication of a metacognitively guided strategy of choosing the lowest JOL items for study, it did not, consistently, improve performance. Apparently, accurate metacognitive monitoring is largely in place in middle childhood, but is not yet converted into effective implementation strategies. This dissociation between metaknowledge and its implementation in choice behavior needs to be taken into account by educators aiming to design interventions to enhance learning in children at this age.

Keywords Metacognitive monitoring and control · Children · Development

J. Metcalfe (✉)
Department of Psychology, Columbia University, New York, NY 10027, USA
e-mail: jm348@columbia.edu

B. Finn
Department of Psychology, Washington University in St. Louis, St. Louis, USA
e-mail: bridgid.finn@wustl.edu

Going back at least as far as Woodworth (1921) it has been thought that monitoring and control skills are essential in allowing both children and adults to optimally devise their own study plans, and to engage in self-guided learning in an efficient and sustained manner. Nelson and Narens (1990, and see Schraw and Moshman 1995) delineated a broad theoretical framework whereby people's metacognitions, that is, their knowledge of their own knowledge, could be used to control their own study and enhance learning. Within this framework, good metacognition is essential for control of learning (Benjamin et al. 1998; Dunlosky and Hertzog 1998; Koriat 2002; Mazzoni and Cornoldi 1993; Metcalfe 2002; Nelson and Dunlosky 1991; Pressley and Ghatala 1990; Reder 1988; Thiede 1999). Of course, much learning, both in animals and in humans, is either driven directly by the environment, by the promise of rewards, or by habitual default strategies (see, e.g., Ariel et al. 2009; Dunlosky and Ariel 2011; Metcalfe and Jacobs 2010) and is not metacognitively guided. But, in some cases, as Nelson and Narens (1990) proposed, learners can, actively, take control of their own learning. Effective internally-driven self-regulated learning is a highly prized goal of education.

Two explicit models of such metacognitively-guided self-regulated learning have been proposed: The Discrepancy Reduction model (Dunlosky and Hertzog 1998; and see, Dunlosky and Thiede 1998; Thiede 1999) and the Region of Proximal Learning model (Metcalfe 2002; Metcalfe and Kornell 2003, 2005, which draws on Atkinson 1972; Berlyne 1978; Hebb 1949; Vygotsky 1987). Both of these models depend upon learners using their metacognitive knowledge to determine their study choices and time allocation, and, in both models good metacognitive knowledge is central. Furthermore, both models agree that it is inefficient for people to study items that they already know they know well. If an item is already mastered, both models agree that people should decline further study. The two models diverge in what people should do once the already-known items have been eliminated from contention. The Discrepancy Reduction Model says that the learner should allocate their efforts to the most difficult items, reducing the biggest discrepancies from the learning state first and selectively. This means that the Discrepancy Reduction Model predicts that the gamma correlation between judgments of learning and study choice should be highly negative, if the person is behaving strategically. In contrast, the Region of Proximal Learning model says that people should devote their efforts to the easiest as-yet-unlearned items, picking off the 'low hanging fruit' for study—fruit that is in the zone, specific to each individual, of what can readily be mastered. Only then should they turn to more difficult items. This means that sometimes gamma correlations between judgments of learning and study choice may not be highly negative, especially when few already-learned items are included in the choice set. Although these two models differ (in testable ways) on how they propose the learner should use their metacognitive knowledge to optimize study, they agree that (1) it is essential that the person be able to make accurate metacognitive judgments which become the basis for choice, and (2) this metaknowledge must be converted into an effective strategy that the learner then implements. If either of these components is weak, then self-regulation of study will also be weak.

It seems self-evident that children, like adults, need to realistically assess their own knowledge, to allocate and organize study time and effort optimally, and to do so with a diligence and verve that will be rewarding and result in enhanced learning and provide the basis for an appreciation of and a taste for learning. Surprisingly, although there has been great progress in understanding adult metacognition and its use in study choice and time allocation, and although there are some experimental studies with children dating back decades (Dufresne and Kobasigawa 1989; Masur et al. 1973), there are only a handful of recent experimental studies (Lockl and Schneider 2002; Son 2005, and see, Arnold 2009; de Bruin et al. 2011) investigating children's metacognitive capabilities and their translation into study strategies.

In this article, then, we investigate the development of young children's metacognitive and control ability at the Grade 3 and 5 levels. This is the age at which school children are starting to need to actively control their own learning. The number of hours that children at the elementary school level must spend on homework, that is, in unsupervised study, has increased dramatically (Hofferth and Sandberg 2000), and increasing pressure is being placed on children to produce results. However, as Cooper et al. (2006) noted the beneficial effects of homework vary considerably across grades. Though the hours that students in high school spend on homework have a sizable effect on achievement, there is a much smaller effect of homework at the junior high level, and there is *no benefit* to homework and self-study at the elementary school level: "The correlation between time spent on homework and achievement was not significantly different from zero for elementary school children" (Cooper et al. 2006, p., 43). Although we recognize that the connection is still speculative and other causes for the lack of effectiveness of homework at a young age might also be important, one possible reason for the lack of effect of the hours spent on homework among the elementary school children may be that either their study strategies or their metacognition is faulty. They may not know what they know or do not know. If they do have adequate metacognitive knowledge they may not put that knowledge to use effectively in implementing study strategies that enhance their learning. So, they may truly labor in vain because of metacognitive/control immaturity. Understanding what children's metacognitive/control capabilities are at this age is both theoretically important for understanding their cognitive development, and it may also have practical importance in allowing educators to set challenges and offer strategies (or, perhaps to simply offer more direct scaffolding) that can help the children to learn more effectively.

We investigated two hypotheses concerning children's metacognition and control, both of which have some, but not unequivocal, support in the extant literature. The first hypothesis was that elementary-aged children's metacognitive monitoring ability may not be fully developed—they may have poor resolution or relative accuracy, not knowing which items they know better and which they know worse. Second, it is possible that even with adequate metacognitive knowledge, they may not convert that knowledge into effective study strategies.

Many studies have investigated metacognitive knowledge by asking people to provide judgments of learning or judgments of knowing about the to-be-learned materials, and then administering a test on those materials. To the extent that the metacognitive judgments are correlated with performance (often using an item by item gamma correlation, see, Nelson 1984) we may assess the relative accuracy or resolution of the metacognitions. These judgments also allow investigation of whether people are over or under confident—thinking they know more or sometimes less than they really know as demonstrated by their test behavior. But even if people's knowledge of their knowledge is accurate, they may still choose inappropriately.

One method of assessing their choices is to allow them to choose what they wish to study, then either honor or dishonor those choices (Kornell 2005). If performance is better when their choices are honored, then they have chosen well. A number of studies have shown that adults' performance is enhanced when they are given the items to study that they chose rather than the items that they declined to study (Kornell 2005; Kornell and Metcalfe 2006; Nelson et al. 1994; Son 2010). Alternatively, one can provide participants with items to study based on some presumably optimal model (Atkinson 1972; Metcalfe et al. 2007) and compare the learning result to a condition in which the participants made their own study choices. These methods will be used to investigate whether children exhibit what Schneider and Pressley (1997) have called an 'implementation deficit.'

There are some indications in the literature that children's knowledge about their own knowledge may not be fully developed at the Grades 3–5 level. Flavell et al. (1970) showed that older children were better able to predict their memory span than were younger children. Markmann (1977) showed that younger children were less able to pick up on inconsistencies and comprehension errors than were older children. Brown et al. (1983) showed that the effective selection of cues developed right up to college-age. Wellman (1977) showed an increase in metacognition with age, but only tested kindergarten to Grade 3 children. Wellman (1978) showed that 10 year olds but not 5 year olds could tell the relevance of two memory variables interacting, such as the number of items, retention interval, or if the memorizer was fresh or fatigued, thus illustrating more sophisticated metacognition. In contrast, Wellman et al. (1981) found that even children in kindergarten could take account of two memory variables, and showed good metacognitive monitoring. Pressley et al. (1984) showed that there was a developmental trend in the grasp of mnemonic strategies between 11 year olds, 13 year olds, and adults. All needed experience, but the younger children also needed explicit feedback. Pressley et al. (1987) found that even young children exhibited good predictive ability, but that the predictions got better still with age. Lockl and Schneider (2002) found that older children were better than the younger children on a feeling of knowing task. Monroe and Lange (1977) and Bisanz et al. (1978) also found metacognitive improvements with age.

Whereas both adults and children tend to be overconfident, particularly on the first study trial (see, e.g. Koriat et al. 2002; Koriat & Shitzer-Reichert 2002), several studies also suggest that younger children may be more overconfident than older children. Worden and Sladewski-Awig (1982) found that the youngest children (kindergarten and Grade 2) knew equally as well as older children (Grade 6) what they knew when they did in fact know it. The younger children were, nevertheless, more likely to think that they knew things they did not than were older children. Such a wishful thinking bias (Schneider and Pressley 1997) in which one believes one knows what one does not yet know could have a deleterious impact on the choice of items to study. For example, the Region of Proximal Learning model (Kornell and Metcalfe 2006; Metcalfe 2009, 2010; Metcalfe and Jacobs 2010; Metcalfe and Kornell 2003, 2005) proposes that these ‘almost but not quite learned’ items are the best candidates for study, and most open to learning with the least time and effort. If one falsely believed one knew these items and therefore declined studying them further, one would miss studying the items most likely to benefit from that additional effort.

In the experiments that follow we assessed the children’s metacognition both by having them use an analogue scale which allowed them to indicate that they had partial knowledge, and also by making a binary classification into items they thought they knew and those they thought they did not know. We expected, as the above mentioned studies suggest, that overconfidence might be greater with the grade 3 than with the grade 5 children. We also expected that, regardless of age, the binary judgments of knowing might reduce overconfidence.

Although there are suggestions that children may have a metacognitive deficit at this age not all studies are consistent with this conclusion. Schneider et al. (2000) found no effect of age (from Kindergarten to Grade 4) in two experiments that compared immediate and delayed judgments of learning (JOLs). Roebers et al. (2007) investigated children’s monitoring abilities in a complex, everyday memory task. They found no differences in the accuracy in a JOL task between 8 year olds, 10 year olds, and adults. Cultice et al. (1983) and Kelly et al. (1976), found that young children exhibited good metacognitive abilities. Butterfield et al. (1988) found that 1st graders were actually better than older children on a feeling of knowing task. So, there is some indication that basic monitoring may be in place by Grade 3. Nevertheless, Bjorklund and Douglas’ review (1997, p. 221) reached the conclusion that “in general, research over the past two decades reveals that metamemory knowledge increases with age.”

One possibility for the inconsistent results concerning the development of children’s metacognitive ability, made salient by investigations of the mechanisms underlying adult metacognitive judgments, is that the accuracy of children’s metacognitive judgments may depend, at least in part, on how the experimenter elicits the judgments. Even in adults, there are some metacognitive tasks that consistently result in illusions of metacognition, in inaccurate judgments, and in biases. But not all do. For example, there is now an extensive literature that shows that delayed JOLs (in which only the cue, but not the target, is presented) result in much higher correlations between the predictive judgments and later test performance than do immediate judgments (Connor et al. 1997; Dunlosky and Nelson 1994; Kelemen and Weaver 1997; Weaver and Keleman 1997). Typically, in adults, the correlation between the judgment and performance is of the order of 0.35 with immediate judgments, but often exceeds 0.85 with delayed judgments. Furthermore, the calibration of the JOLs, that is, how closely the average judgments match the average level of performance, is better with delayed than with immediate JOLs. In the few studies that have investigated delayed JOLs in children, those judgments have been fairly accurate (Roebers et al. 2007; Schneider et al. 2000; Koriat and Shitzer-Reichert 2002). Few target materials have been investigated in this way, however.

We thought it would be inappropriate to test children using the immediate JOL paradigm, or in a metacomprehension paradigm, in which even adults cannot consistently make accurate metacognitive judgments, and in which knowledge of heuristics that could promote accuracy may not be readily available to adults and especially not to children. Both for practical reasons, and to avoid drawing conclusions that may underestimate the children's abilities, we thought it best to provide a situation in which we could optimize the chance of them showing their metacognitive capabilities, if such capabilities existed. Thus, in the first three experiments, we used a delayed JOL task, and a variety of materials: vocabulary, Spanish-English translations (paired-associate form), and general information questions.

The last three experiments address the question of the efficacy of children's choices of what to study. In adults, there is a connection between metacognitive judgments of learning and study choices (Finn 2008; Metcalfe and Finn 2008; Thiede et al. 2003). Manipulations of the judgments result in changes in study choice. Furthermore, adults often choose to study what they think they do not know (Dunlosky and Hertzog 1997; Nelson et al. 1994; and see Dunlosky and Ariel 2011; Son and Metcalfe 2000, for reviews). There is some indication that children may not choose in the same way as adults, however (e.g. Bisanz et al. 1978; Dufresne and Kobasigawa 1989; Flavell et al. 1970; Masur et al. 1973). For example, Bisanz et al. (1978) found that whereas Grade 5 and college students selected unrecalled items for study, Grade 3 and Grade 1 children did not. Similarly, we will investigate whether children eliminate items they know they know and choose the items they know they do not know to study. We will also compare children's choices to those predicted by the Region of Proximal Learning and the Discrepancy Reduction model. When the already-known items are eliminated, do children choose to study those items that are closest to being learned as the Region of Proximal Learning model would predict, or do they select the items that they consider most difficult, as the Discrepancy Reduction model would predict? And do these choices enhance performance? We will address this question by using the honor/dishonor methodology. These last three experiments allow investigation of the possibility that children at this age, even with adequate metacognitive knowledge, may have an implementation deficit.

Our participants were Grade 3 and Grade 5 children. We chose these age groups because they span the range where, as noted above, we expected to see changes in the development of the metacognitive and choice capabilities. By college, these capabilities are maximally in place, and hence this group provides a benchmark for the younger children. We know that college students have highly accurate metacognitions, and that they usually put these to use in making study choices. Further, honoring their choices enhances their learning (Kornell and Metcalfe 2006; Nelson et al. 1994; Son 2005). The first three experiments presented here investigate children's metacognition; the last three experiments look at their choices and the relation of these choices to their metacognitions and their learning in a paradigm in which adults have been shown to exhibit enhanced learning when their choices are honored. The hypotheses for the first three experiments, to which we now turn, were that (1) the relative accuracy of younger children's metacognitive judgments, indexed by the gamma correlation between their JOLs and recall performance on individual items, would be worse than that of the older children (and neither would be as good as is typically seen in adults), and that (2) the younger children would be more overconfident than the older children.

Experiment 1

The objective of the first experiment was to investigate whether children know what they know and don't know. To determine both the relative accuracy of JOLs (as given by gamma correlations), and the cutoff between items that the children thought they knew and did not know, we asked for two judgments in this two-session experiment. During the first session, the children made JOLs on a scale from 0 to 100 for each item, and took a criterion test. In the second session, the children decided whether they knew or did not know each item (JOKs). These measures allowed us to compute relative accuracy gammas, as well as assess over or underconfidence.

Method

Participants The participants were 21 children, 13 from Grade 5 (M_{age}=10.00 years) and 8 from Grade 3 (M_{age}=7.75 years), from the Emily Dickinson School, PS-75, in New York City. Approximately half the children in each group were girls. Only children who spoke fluent English were included in our sample. The children participated in an after-school program at Barnard College once a week, on three different days. There was no financial compensation for participation but the children were always given snacks before each day's activities, took part in interesting group activities when they were not participating in the experiment, and participated in a 'Columbia-Barnard party' at the end of the year. In this and in the studies that follow, the treatment of participants was in accordance with APA ethical standards. The sample was racially diverse (13 % White, 28 % Black, 51 % Hispanic and 7 % Asian). Although precise information about the sample's SES was not available, the overall poverty rate of the PS-75 school population was 67 %. Participant recruitment, the sample's gender, mean age, ethnicity and socioeconomic backgrounds were the same in all of the experiments that follow.

Design The experiment consisted of 2 Judgment Types: JOL or JOK, which were assigned such that on the first session the children had only JOLs, whereas on the second they did both JOLs and JOKs, within-participants. Grade (3 or 5) was a between participants variable.

Materials The materials were 24 definition-word pairs drawn from school textbooks and online vocabulary resources. There were different lists for the two grades in order to attempt to match their level of difficulty. The materials for all our experiments are available on request.

Procedure The experiment was conducted on iBook computers. Each child was tested individually, in a sound buffered room, with an experimenter/coach. The experiment proper was embedded in what we call the "Dragon Master" shell—a rewarding game-like program that we developed to make studying facts and other experiments testing the children more engaging. The "Dragon Master" is an imaginary wizard who encourages the children to learn well so they can earn gold and progress from being an egg to a full-blown dragon, see Metcalfe et al. (2009) and Metcalfe and Finn (2008) for more details.

Before the experiment started, the students went through a practice trial to ensure that they understood what JOL was and how to use the slider scale in the program. They were given 3 easy (What is your name?) and 3 impossibly difficult (What is the experimenter's middle name?) questions to make JOLs on and through which they learned how to use the high and low ends of the JOL scale. They were also given three questions that had ambiguous answers, to teach them how to use the middle of the scale. One example of such a question was, "Will it rain tomorrow?" The practice session was repeated if the child had not understood all three types of questions well, or needed more practice controlling the JOL slider.

The experiment consisted of two sessions. In the first session, the participant studied 12 definition-word pairs. Each definition appeared on the screen and was followed by the correct word. A recorded voice read aloud the definition, repeated the word twice and read a context sentence in between the two repetitions. The participant chose when s/he was ready to move on to the next word. This process was repeated for 12 pairs. The participant was then told that s/he would be making JOLs for each word that he or she had studied. For each JOL, the definition appeared on the screen and the participant had to move a slider to indicate how well he or she had learned the answer. After 12 JOLs, participants were tested on all 12 studied items. The definition appeared on the screen and was also read out aloud and the participant had to type in the correct word. The program had a built in formula that calculated a score for the response based on the way it was spelled (and which, in effect, ignored spelling errors). A score of 75 and above was treated as correct (and corresponds to what adult scorers would say are just spelling errors or typos, and hence are the correct answer) and the participant was rewarded with a piece of "gold" for getting each answer correct. Corrective feedback was given after each response.

The study phase of the second session was similar to the first. In the JOL phase, the participant first made a JOL, which was immediately followed by a Judgment of Knowing (JOK) for each item. The JOK was a Know/Don't Know judgment in which the participant had to click one of two buttons. The test on the definitions was given after all judgments were made, and was identical to that of session 1. The order of the items was randomly selected in the study, JOL/JOK and test phases in each session, and the order of the sessions was counterbalanced.

Results

The grade 5 children (0.34) performed equivalently to the Grade 3 children (0.23), $t(19)=1.19$, $p>.05$.

We used gamma correlations to measure relative accuracy. This measure, as advocated by Nelson and Narens 1990, has been used in most previous studies on relative accuracy (and thus the results of those studies can be compared to ours) and seems especially appropriate when there are few ties in the data (see Benjamin and Diaz 2008) as was the case when a visual analogue slider judgment scale is used in our studies. The relative accuracy scores—indexed by the gamma correlations between the JOLs with whether or not the items were later correct on the recall test—were high for both age groups (mean $\gamma=.90$ for the Grade 3 children, and $\gamma=.79$, for the Grade 5 group)—indicating good metacognition. There was no difference between the two groups, $t(17)=1.28$, $p>.05$. Correlations were also computed for the know/don't know judgments. Like the JOL gammas, they were high for both grades, which were not different from one another ($t<1$, $p>.05$). See Fig. 1a.

Both groups showed overconfidence, such that their JOLs were higher than was performance, $F(1,19)=45.26$, $MSe=.91$, $p<.01$, $ES=.70$, $P=1.00$ (observed power will be reported as P throughout), as is shown in Table 1. The interaction with grade, was not significant, $F(1,19)=1.28$, $p>.05$, $P=.19$, nor was there a significant main effect of grade, $F<1$, $p>.05$,

P=.12. As is also shown in the table, asking the children whether they knew or did not know the answers, somewhat offset their overconfidence. The difference between JOLs and JOKs, was significant, $F(1,17)=7.53$, $MSe=.09$, $p=.01$, $ES=.31$, $P=.74$ but, again, the interaction between JOLs versus JOKs and grade was not significant ($F(1,17)=2.25$, $p>.05$, $P=.29$).

We also divided the data into those questions that children got right, and those that they got wrong, and looked at the corresponding metacognitive ratings. The difference in overconfidence between the two kinds of judgments was particularly apparent in the answers that the children got wrong on the test—which showed higher mean metacognitive judgments when the children gave JOLs than when they gave JOKs (Table 1). The interaction between judgment type (JOL versus JOK) and whether or not the answers were correct or incorrect, was significant, $F(1,14)=11.75$, $MSe=.13$, $p<.01$, $ES=.46$, $P=.29$. It would seem that when people made JOLs but not JOKs, they took partial learning into account. This was supported by post hoc comparisons demonstrating that the difference between JOLs and JOKs for incorrect items was significant, $t(18)=3.16$, $p<.01$, whereas there was no corresponding difference shown between JOKs and JOKs for correct items ($t(15)=1.03$, $p>.05$).

Finally, we looked for a correlation between high relative accuracy (i.e., good metacognition) and performance. It seems intuitive that there might be such a correlation—that good metacognition could lead to good strategies of learning, or even that children who are good at doing metacognitive tasks are the more advanced children who might be expected to do better on the recall tests. Although many researchers have suggested that this might be so, few have actually found the effect (Kao et al. 2005; Maki and Berry 1984), and many have failed to find it (e.g., Begg et al. 1992; Cavanaugh and Perlmutter 1982; Kelly et al. 1976). It was not the case, in these data, that the children who had the highest performance scores were also the children who had the best metacognitions, as measured by their JOL to performance gammas. The Pearson correlation relating the metacognition gamma to performance, across children, was only 0.02, and was not significantly different from zero, $p>.05$.

Experiment 2

The second and third experiments were replications of the first experiment using different materials. These allow us to investigate the generality and limitations of

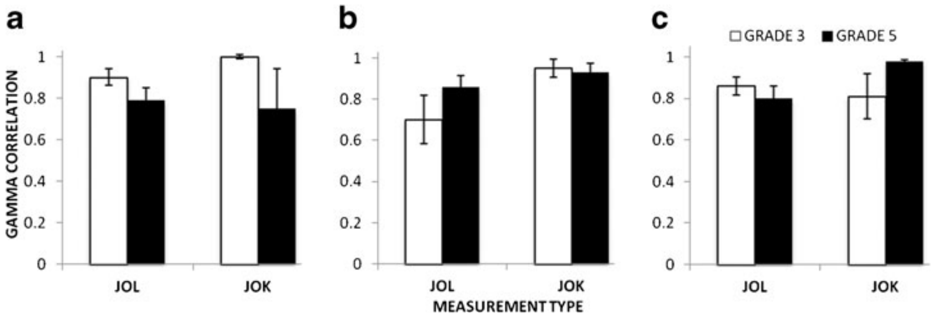

Fig. 1 Metacognitive relative accuracy (gammas between accuracy and JOL or JOK) in Grades 3 and 5 in Experiment 1 with vocabulary (**a**), in Experiment 2, with Spanish-English pairs (**b**), and in Experiment 3, with General Knowledge questions (**c**). Error bars indicate the standard errors of the mean

Table 1 Mean Recall, JOLs, JOKs and JOLs and JOKs on Correct and Incorrect items, for Experiment 1, with vocabulary, Experiment 2, with Spanish-English pairs and Experiment 3, with General Knowledge questions

	Measure						
	Recall	JOL			JOK		
	Overall	*Overall*	*Incorrect*	*Correct*	*Overall*	*Incorrect*	*Correct*
Experiment 1							
Grade 3	.23 (.05)	.59 (.04)	.42 (.07)	.94 (.10)	.42 (.08)	.20 (.11)	1.0 (.12)
Grade 5	.34 (.06)	.59 (.05)	.42 (.05)	.85 (.07)	.54 (.07)	.32 (.07)	.85 (.08)
Experiment 2							
Grade 3	.61 (.04)	.67 (.04)	.37 (.08)	.86 (.04)	.61 (.02)	.09 (.07)	.84 (.04)
Grade 5	.48 (.05)	.58 (.04)	.29 (.05)	.89 (.03)	.56 (.04)	.22 (.05)	.92 (.03)
Experiment 3							
Grade 3	.38 (.05)	.57 (.04)	.35 (.06)	.90 (.03)	.52 (.05)	.33 (.08)	.89 (.04)
Grade 5	.64 (.05)	.73 (.05)	.44 (.05)	.91 (.03)	.70 (.06)	.35 (.07)	.97 (.04)

Numbers in parentheses indicate standard error of the m

the metacognitive effects seen in the first experiment. In Experiment 2 we used Spanish-English vocabulary pairs.

Method

Participants The participants were 22 children, 15 from Grade 5 and 7 from Grade 3, from the Emily Dickinson School in New York City. Some of the participants in this experiment may have also participated in Experiment 1 or 3, but we could not isolate individuals because participant names were not associated with their data.

Materials The materials in this experiment were 24 Spanish-English word pairs of varying difficulty. Grade 5 participants were given 50 % easy words, 30 % medium-difficulty words and 20 % difficult words. Grade 3 participants were given 50 % easy words and 50 % medium-difficulty words.

Procedure The procedure for this experiment was similar to that of Experiment 1, except that the materials were Spanish-English pairs. In the judgment phases and the test, the participants were shown the Spanish word and asked to enter the appropriate response (a JOL, JOK or, in the case of the test, the corresponding English word). There were no context sentences in this experiment. A recorded voice on the computer read out both the words while they were also being presented visually during study.

Results

The Grade 5 children performed numerically worse (0.48) than did the Grade 3 children (0.61) on the recall test, (note that the materials were easier for the Grade 3 group) but this difference did not reach significance ($t(20)=1.64$, *n.s.*). The relative accuracy scores, that is, the gamma correlations between the JOLs and whether or not the items were later correct on the recall test were again, high for both age groups, (mean $\gamma=.70$ for the Grade 3 children, and γ=.86, for the Grade 5 group)—indicating good metacognition. There was no significant difference between gammas of the two groups, $t(20)=1.30$, *n.s.* Correlations were also computed for the know/don't know judgments, and they too were high for both grades, (0.95 and 0.93, for Grades 3 and 5, and not different from one another, t<1) as shown in Fig. 1b.

The amount of overconfidence in this experiment was small, but significant, and is shown in Table 1. The mean JOL for the Grade 3 children was 0.67, and it was 0.58 for the Grade 5 children, for a difference between their accuracy and JOLs of 0.06 and 0.10, $F(1, 20)=6.70$, $MSe=.06$, $p<.05$, $ES=.25$, $P=.69$. The interaction between grade and measure (JOL versus performance accuracy) was not significant, $F<1$, *n.s.*, $P=.10$. The near calibration between the metacognitive ratings and performance was also borne out when we analyzed the children's JOKs in relation to their recall. In this case, the Grade 3 children's JOKs of 0.61 were slightly but not significantly *under* confident. The Grade 5 children's JOKs were overconfident by only 0.08. Overall, overconfidence was not different from zero, $t(20)=1.66$, *n.s.*, and there was no difference between the two groups, $t(19)=1.33$, *n.s.*

As is also shown in Table 1, the JOLs were higher than were the JOKs, especially on the items that were later incorrect. There was a main effect of judgment type, $F(1,19)=9.08$, $MSe=.02$, $p<.01$, $ES=.32$, $P=.82$, and the interaction between judgment type (JOL versus JOK) and whether or not the answers were correct or incorrect, was significant, $F(1,19)=11.96$, $MSe=.16$, $p<.01$, $ES=.39$, $P=.91$. As in Experiment 1, post hoc comparisons revealed that the difference between JOLs and JOKs for incorrect items was significant, $t(20)=2.86$, $p=.01$, whereas there was no corresponding difference shown between JOKs and JOKs for correct items ($t<1$, n.s.). Although there was no main effect of grade ($F(1,19)=1.43$, *n.s.*, $P=.21$), there was an interaction between judgment type (JOL versus JOK) and grade, $F(1,19)=5.40$, $MSe=.08$, $p<.05$, $ES=.22$, $P=.60$, though none of the post-hoc tests were significant, (largest $t(19)=1.90$, *n.s.*). When the children in both Grade 3 and 5 thought they knew something, their probability of being correct was high ($M=.85$, $SE=.05$), and when they thought they did not know it, their probability of being correct was low ($M=.21$, $SE=.04$, $F(1,19)=150.79$, $MSe=.03$, $p<.001$, $ES=.88$, $P=1.00$). There was a main effect of grade. Grade 3 children showed a larger bias than Grade 5 children, $F(1,19)=6.32$, $MSe=.27$, $p<.05$, $ES=.25$, $P=.67$.

Finally, the children who had the highest performance scores were not necessarily the children who had the best metacognitions, as measured by the correlation between metacognition gammas and performance across individuals, $r=-.002$, *n.s.*

Experiment 3

This experiment was like Experiment 1 and 2, except that it used general information questions as the materials.

Method

Participants The participants were 24 children, 13 from Grade 5 and 11 from Grade 3, from the Emily Dickinson School in New York City.

Materials The materials in this experiment were 24 general information questions drawn from school textbooks and grade-appropriate online resources. Both grades received the same list of questions in this experiment.

Procedure The procedure for this experiment was the same as Experiment 1. In this experiment, there were no recordings of the questions but the research assistants would read out the question if the child had any difficulty. During the test, the participant was shown the general information question and asked to type in the answer.

Results

The Grade 5 children performed better (0.64) than did the Grade 3 children (0.38, $t(22)=3.45$, $p<.01$). This is not surprising insofar as the materials were the same for the two grades. The relative accuracy scores, that is, the gamma correlations between the ranking of the items by their metacognition with whether or not the items were later correct on the recall test were high for both age groups, both for the JOLs ($M=.86$ for Grade 3, and $M=.80$ for Grade 5) and for the know/don't know judgments ($M=.81$ for Grade 3, $M=.98$ for Grade 5), as shown in Fig. 1c. There was no effect of grade ($F(1,20)=1.00$, *n.s.*, $P=.16$), or judgment type, and the interaction between grade and judgment type was not significant.

Children in both grades were overconfident. When JOL and accuracy were treated as independent variables, there was a significant main effect of measure, indicating overconfidence, $F(1,22)=40.47$, $MSe=.23$, $p<.01$, $ES=.65$, $P=1.00$. The interaction between measure and grade was also significant, $F(1,22)=4.53$, $MSe=.03$, $p=.05$, $ES=.17$, $P=.53$, with the Grade 3 children being more overconfident. The mean JOL for the Grade 3 children was 0.57, and it was 0.73 for the Grade 5 children, for a difference between their accuracy and JOLs of 0.19 and 0.09, respectively, $t(22)=2.13$, $p=.05$, as is shown in Table 1.

Treating JOK and accuracy as independent variables, there was a significant main effect of measure indicating overconfidence, $F(1,22)=55.25$, $MSe=.12$, $p<.01$, $ES=.72$ $P=1.00$. The Grade 3 children were more overconfident than the Grade 5 children, $F(1,22)=8.33$, $MSe=.57$, $p<.01$, $ES=.28$, $P=.78$. The interaction between measure and grade was also significant, $F(1,22)=6.70$, $MSe=.01$, $p<.05$, $ES=.23$, $P=.70$. The difference between the Grade 3 JOK scores and recall accuracy and the Grade 5 JOK scores and recall accuracy showed that the former were more overconfident than the latter, $t(22)=2.59$, $p<.05$. In sum, both the JOLs and the JOKs showed overconfidence and both showed more overconfidence for the Grade 3 children than for the Grade 5 children.

Although there was a slight numerical difference between the two kinds of judgments in the answers that the children got wrong on the test—which were higher when the children gave JOLs than when they gave JOKs—the interaction between judgment type (JOL versus JOK) and whether or not the answers were correct or incorrect, was not significant. There was, of course, an effect of whether the response was correct or incorrect on these judgments ($F(1,22)=142.11$, $MSe=.05$, $p<.001$, $ES=.87$, $P=1.00$), but no other effects or interactions were significant. These data are shown in Table 1.

When children of both grades said they did not know something their probability of being correct on those items was low ($M=.06$, $SE=.03$). When the Grade 5 children said that they knew the answer, their performance was high ($M=.89$, $SE=.02$). But when the Grade 3 children said that they knew, their probability of being correct was only slightly better than 50 %. The interaction between measure and grade was significant, $F(1,20)=16.85$, $MSe=.40$, $p<.01$, $ES=.46$, $P=.97$. There was no difference between the two grades when they said they did not know the answer, $t<1$, but Grade 3 children's probability of being correct was significantly lower than that of the Grade 5 children when they claimed that they knew it, $t(20)=5.02$, $p<.01$, indicating that the younger children, in this experiment, were engaging in something that might be called wishful thinking.

There was no relation between metacognition gamma and memory performance across children, r=.015, *n.s.*

Discussion of experiments 1, 2, and 3

None of the three experiments indicated that the younger children's metacognitive relative accuracy was worse than that of the older children. Indeed the relative metacognitive accuracy of both the Grade 3 and the Grade 5 children was as high as that typically seen with adults. All of the children ranked the items that they would eventually be able to recall higher than those they would not, with high accuracy. Thus, the first hypothesis, that elementary-aged children's metacognitive accuracy would show poor resolution, appears to be incorrect, at least when delayed cue-only JOLs are used as the metacognitive task.

The second hypothesis was that younger children would be more overconfident, or exhibit more wishful thinking than older children. Whereas Experiment 3 supported the hypothesis, Experiment 1 showed no differences between the two age groups, and Experiment 2 went in the opposite direction: the older children were more overconfident than the younger children. Furthermore, when JOKs were used as the measure of confidence rather than JOLs, the extent of overconfidence decreased. Furthermore, when the younger children did show more overconfidence than the older children, in Experiment 3, they also exhibited worse performance. Studies with adults show a relation between poor task performance levels and overconfidence (Connor et al. 1997; Hacker et al. 2010; Kruger and Dunning 1999; Richards and Nelson 2004; Scheck and Nelson 2005), making it difficult to interpret the selective overconfidence in Experiment 2 as being attributable to age differences alone. When performance levels were higher and close to being the same in the two groups, in Experiment 2, overconfidence was smaller and the same in the two groups. Neither Experiment 1 nor 2 showed a difference in overconfidence between the older and younger children. This pattern of results, overall, then, offers little support for the idea that the younger children were more overconfident, especially when task performance levels were about the same.

Finally, in none of the three experiments did we observe a relation between good performance overall, and good metacognition as measured by the relative accuracy gammas. The correlations, in all cases, were not significantly different from zero. Thiede et al. (2003) have noted that having the knowledge of what you do or do not know may not help if there is no way to use that knowledge to study further or to improve one's performance. The next three experiments will investigate the children's use or perhaps lack of use of what is good metacognitive knowledge, by investigating their restudy choices.

Experiment 4

The hypothesis that we tested in Experiment 4 was that although there did not appear to be any impairment in children's metacognitive knowledge, as given by their relative accuracy, there might, nevertheless, be developmental differences in how effectively children converted their (excellent) metaknowledge into viable study choice strategies. There are several indications of this possibility in the literature. Bisanz et al. (1978), for example, found that whereas Grade 5 and college students selected unrecalled items for study, Grade 3 and Grade 1 children did not. Although there are differences among theories as to exactly what comprises an optimal choice strategy, all theories (including those advocating the discrepancy reduction rule, such as Dunlosky and Thiede 1998, 2004 and those proposing that the person should try to study in their own Region of Proximal Learning, e.g., Metcalfe 2002, 2009, 2010; Metcalfe and Jacobs 2010; Metcalfe and Kornell 2003, 2005) agree that people should not choose to study the items that they have already firmly mastered—the idea that we tested in this experiment. And, although typical adults (see, Knouse et al. 2013, for contrasting strategies used by adults with ADHD) will sometimes also use habitual default strategy (such as reading from left to right, see, Ariel et al. 2009; Dunlosky et al. 2011) they generally avoid studying already-known items. Insofar as the young children, in Bisanz et al. (1978), apparently did not eliminate these items, it seems reasonable to hypothesize that they did not convert their metacognitions to optimal choices.

There are several other studies, as well, that suggest that younger children may choose differently or allocate study time differently than do older children. For example, Masur et al. (1973) found that first grade children were less likely to choose unrecalled items for restudy than both Grade 3 children and college students. Dufresne and Kobasigawa (1989) also found that younger children spent the same amount of time studying easy and difficult words whereas older children spent more time on the difficult words. As Schneider and Pressley (1997, p. 217) put it: “apparently knowing which information is known already or easier to learn and which information is unlikely to have been mastered is not sufficient to result in appropriate self-regulation.”

Though such findings might indicate that the younger children were choosing inappropriately, it is also possible that the easier items they chose were, in fact, in the children’s Region of Proximal Learning and they were choosing an optimal strategy. Even restudying the items that they had been able to recall might have helped performance of the younger children. Perhaps these items were newly learned and fragile and needed more study to be firmly consolidated. If this were so then it is possible that even though the younger children did not choose the most difficult items for study—unlike adults and older children—both the younger and the older children might have been choosing the items that were right for them. Furthermore, if one were to examine the children’s choice distributions, they may have shown a preference for the items of middle difficulty, rather than simply being random across the JOL range. They may still have eliminated the *very* well learned items, but, perhaps, not have preferred the most difficult items. Son (2005) has shown that this non-monotonic choice prediction sometimes holds up with children.

The correlation between JOL and choice has often been used to indicate differences in strategy, and differences in this correlation may be important. We will look for differences in this correlation in the study that follows. However, a lower JOL to choice correlation in younger children might also result from younger children correctly and strategically choosing to restudy the less difficult, rather than the most difficult, items—a selection that could be right for them. Rather than just looking at what items children chose, and making assumption about efficacy or the lack of efficacy of those choices, we will also evaluate the goodness of the children’s choices with the honor/dishonor procedure (Kornell 2005; Kornell and Metcalfe 2006). In this procedure, participants choose some of the items for restudy. Then the computer either honors the participants’ restudy choices or dishonors them, by giving them for restudy the items that they had declined. The logic is that if people do better on the final test after restudying items they had chosen themselves, as compared to those they had declined, then they had chosen the right items for them. Thus, this honor/dishonor paradigm allows stronger inferences about the efficacy of choice than can be made by investigating only the correlations between the children's choices and their JOLs. In this experiment and the two that follow, then, we sought to determine both whether study choice differed between our Grade 3 and 5 groups, as well as whether the choices were effective: did honoring the children's choices result in better overall performance?

Method

Participants The participants in this experiment were 30 children, 16 from Grade 5 and 14 from Grade 3, from the Emily Dickinson School in New York City, who had not participated in previous experiments.

Design The experiment was a 2 (Condition: Honor, Dishonor) by 2 (Restudied, not Restudied) within-participants design with grade as a between subjects variable. A high JOL condition, was also included. However, it is not relevant for the present hypotheses and is, therefore, not included in the analyses below.

Materials The materials in this experiment were 54 definition-word pairs drawn from school textbooks and vocabulary resources online. Each grade was given a different list of words in an effort to match the level of difficulty across grades.

Procedure The experiment was divided into 3 sessions. The sessions were separated by a week and therefore the whole experiment took 3 weeks to conduct. The experiment was embedded in the Dragon Master shell that was described in the first experiment and the children continued to new levels across the 3 weeks.

The first three sessions were identical and were separated only because of time constraints. The first step was a preliminary test that was conducted to exclude questions to which the children might already have known the answers. They were rewarded at this stage for correct answers and we included a few deliberately simple questions to motivate the children. This test continued until we had a list of 18 items that each child did not know. After the first test, there was a timed study phase during which the child studied 6 items for 5 s each. After the study phase, he or she made JOLs on these six items. After the JOL phase, the participants saw all six questions in a circular arrangement and were asked to choose three to restudy. The study, JOL and choice phases were repeated twice more until 18 items had been studied and offered for choice.

The program randomly assigned each of the three blocks of six questions to one of the three conditions—honor, dishonor and High JOL. In the honor block, the 3 items that the participant had chosen were given for restudy. In the dishonor block, the other 3 out of the 6 items were given for restudy. In the High JOL block, the three highest JOLs were given for restudy. There were a total of 9 items in the restudy phase, which included the given items from all three blocks.

After the restudy, the participant was tested on all 9 items that were restudied plus one item from each condition that was not restudied, for a total of 12 test items. Fewer non restudied than restudied items were given at test because we were concerned that the children might become discouraged when they were tested on many items that, because they were not given restudy opportunity, they did not do well on. No feedback was given at this stage. This whole process was repeated for the first three sessions, with the order of conditions being counterbalanced over sessions. At the end of the third session they had been tested on 36 items.

Results

Metacognition We investigated the goodness of children's metacognitions by using only those items that were not restudied, since the restudy itself, could, of course, alter their learning. The gammas between the children's JOLs and their later recall performance on the non-restudied items were high (0.74 for Grade 3 and 0.77 for Grade 5) and there was no difference between the Grade 3 and the Grade 5 children (t<1).

Choice strategies The Grade 3 children showed near zero gamma correlations (γ=−.09) relating their JOLs to their choices for study, whereas the gammas of the Grade 5 children were highly negative (γ=−.45, t(28)=2.94, p<.01), indicating that the Grade 5 children were being 'more strategic' than were the Grade 3 children. The negative correlation indicated that the Grade 5 children chose for study the items they knew they did not know or that they thought were more difficult or less well learned. The Grade 5 children's gamma correlations between choice and JOL were significantly negative, t(15)=6.77, p<.01. However, the gamma correlations of the Grade 3 children were not different from zero, t<1, *n.s.*, suggesting that they were choosing randomly. Also, as seen in Fig. 2, the mean JOLs of the items that were chosen for restudy were lower than for those items that were declined for restudy for the Grade 5 group (t(15)=5.96, p<.001) but not for the Grade 3 group, t<1, *n.s.* The interaction was significant, F(1,28)=5.73, MSe=.20, p<.05, ES=.17, P=.64. Clearly the strategies used were different for the older than the younger children.

To examine whether the children were making efficacious choices, we computed the mean recall in the honor and dishonor conditions by estimating the total mean recall for that condition based on a weighted average of recall on the 9 restudied items and the 3 non-restudied items (which were weighted equally with the restudied items, i.e., multiplied by 3). The Grade 5 children's means were 0.60 for the honor condition versus 0.52 for the dishonor condition, whereas the Grade 3 children's means were 0.43 honored versus 0.43 dishonored. Neither the main effect of honor/dishonor, F(1,28)=1.42, *n.s.*, nor the interaction between honor/dishonor and grade were statistically significant F(1,28)=1.56, *n.s.* (see Fig. 3). Numerically the Grade 5 children, but not the Grade 3 children benefitted a little from having their choices honored, though that benefit was not statistically significant (t(15)=1.70, *n.s*).

Discussion

Once again, the results of this experiment indicated that the children in both grades had excellent metacognition. Only the Grade 5 children were starting to choose the items that

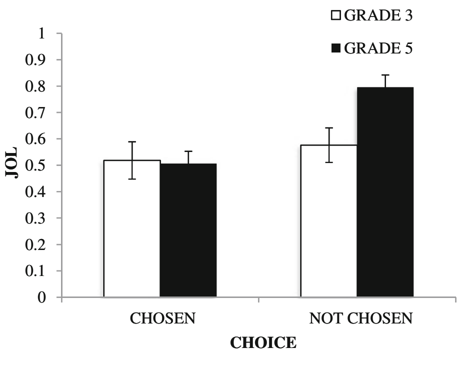

Fig. 2 JOLs of items based on whether they were chosen or not chosen for restudy in Grades 3 and 5 in Experiment 4. Error bars indicate the standard errors of the mean

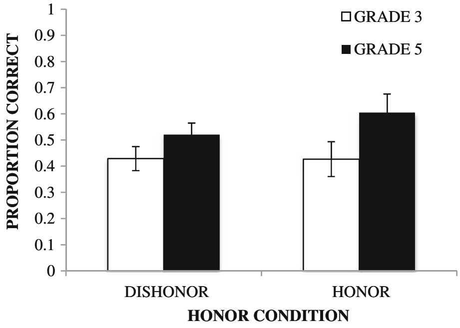

Fig. 3 Mean recall performance in the honor and dishonor conditions in Grades 3 and 5 in Experiment 4. Error bars indicate the standard errors of the mean

they thought they did not know for restudy or that had low judgments of learning. Grade 3 children apparently chose randomly. Honoring the choices of neither the Grade 3 nor the Grade 5 children significantly improved their performance. This pattern is in stark contrast to that seen with adults, in which honoring their choices in this same paradigm results in better recall performance than when their choices were not honored (Kornell and Metcalfe 2006). These results suggest that children at around the Grade 5 level, but not yet at Grade 3, are at the beginning stages of being able to use their metacognitions to make strategic study choices. But even when they are making strategic choices, as indicated by the relation between their metacognition and the choices, these choices are apparently not nuanced enough to result in enhanced performance. Before drawing any strong conclusions, though, we conducted two more experiments.

Experiment 5

Experiment 4 showed that Grade 5 children were choosing more strategically than Grade 3 children. However, the difference in strategy did not result in statistically significant differential performance between the two grades on the honor choice part of the experiment. In the last two experiments we further investigated what appeared to be a failure in the younger children in converting their excellent metacognitive knowledge into any kind of study choice strategy. We also investigate the possibility that the older children are beginning to use a metacognitive study strategy of choosing the least known most difficult items for study. This strategy may be effective under some but not under all circumstances.

As noted earlier, both of the dominant theories of study time allocation, the Discrepancy Reduction model (Dunlosky and Hertzog 1997) and Region of Proximal Learning model (Metcalfe and Kornell 2003, 2005; Kornell and Metcalfe 2006), indicate that people should decline study of the items that they already know. Because this is uncontroversial, this seemed like a clear way to test whether the older children were choosing strategically. It is also the simplest strategy, so perhaps even the younger children might use it, if it were presented in a clear way. In Experiment 5, unlike Experiment 4, we made a sharp distinction between items on which the children knew the answers and were correct and those on which they did not know the answers, and were incorrect (or could not answer). Indeed, the program we used in this experiment was designed to make sure that 12 of the items on which the children would make study choices and be tested were items that they had demonstrated that they already knew (on a pretest) and 12 of them were items that the children did not know (as evidenced by either giving a blank or by giving an incorrect answer on the pretest). To study effectively, they should decline to study the already-known items and selectively choose the items they did not know. If the theories are right, doing so should result in enhanced memory performance.

Method

Participants The participants in this experiment were 23 children, 14 from Grade 3 and 9 from Grade 5, from the Emily Dickenson School in New York City.

Design The design included 4 within-participant conditions: the Honor choice condition, the Dishonor choice condition, the computer designated study Correct item condition and the computer designated study Incorrect item condition. There were children from two grades: Grade 3 and Grade 5. For purposes of data analysis and explication (and because adults do best in this condition), we designate the Honor choice condition a 'good' strategy, and the Dishonor choice condition a 'bad' strategy. Because all theories agree that people should decline study of the already known items (and because this, too, maximizes performance for adult participants) we will also call the computer designated study of Incorrect items a 'good' strategy, and computer designated study of the already correct items a 'bad' strategy. Thus, the design of this experiment can be recast into *self-determined* good or bad strategies (which are the Honor and Dishonor choice conditions, respectively) and *computer designated* good or bad strategy conditions (which are the Correct or Incorrect item conditions, respectively).

Procedure Participants were shown a series of Spanish items and were asked to provide the English translations as the answers. The program continued asking the children for translations until it had amassed 12 questions on which the child got the correct answer and 12 on which the child either left a blank or got the wrong answer. This nearly always meant that there were more questions in one category or the other, and the computer randomly chose 12 questions to be those that were used in that particular condition that had amassed more questions to hit the criterion of at least 12 correct and 12 incorrect. The 12 correct and the 12 incorrect items were then assigned to the four treatment combinations, such that there were 4 choice trials, each of which had three items that had originally been incorrect and three that had originally been correct. In each of the choice trials the children were presented with these 6 cue items, arranged in a circle, and they had to click on the three items for which they chose to be shown the answer to study. They were told that they would get all of the questions on a test in a few minutes, and that they should choose the items for study in such a way that they would get the best possible score on the final test.

The four 6-tuples were assigned randomly to be either in the Honor choice condition, the Dishonor choice condition, the condition in which computer selected for study the items on which the child was originally Correct, or the condition in which the computer selected for study the items that were originally Incorrect. Although the child had no way to distinguish the four choice trials, the order of assignment to condition was counterbalanced over children. Performance scores were assessed for each condition by using the 6 items that were in that 6-tuple and computing the proportion correct on those 6 items.

Once the selection phase was completed, the 12 to-be-studied questions were randomized and presented for study. In the computer determined Correct condition these were the three items that had been correct; in the Incorrect condition, they were the three items that had been incorrect (the theoretically optimal choice), but in the Honor and Dishonor choice, whether the items were correct or incorrect depended on the choices the child made, with the former being the items he or she had chosen, and the latter being the items he or she had not chosen. The child was presented a cue, and asked to type in the correct answer. If the answer the child (or the research assistant, if the child had any difficulty using the computer) typed in was correct, the computer pinged, and left the correct answer onscreen for as long as the child wished to study it. If the answer was incorrect (or the child pressed 'enter' without putting in an answer) the computer presented the correct answer. After the child had studied it for as long as s/he wanted to, and pressed return, the computer then immediately presented that same cue again, and the child had to enter the correct answer, before going on. If they got it wrong a second time the computer presented the correct answer again, they studied again and hit enter. The cue popped up again, and they had to enter the answer correctly, cycling through until entering the correct answer with the correct spelling. Then the next probe was presented, and this cycle continued until all 12 answers were entered, three of which were in the honor choice condition, three of which were in the dishonor choice condition, three of which were in the Correct and three of which were in the Incorrect condition. Then all 24 questions were presented in a random order on the final test.

Results

Learning The mean proportion correct was computed for each of the 4 conditions. As noted above, the Honor condition and the study Incorrect condition were taken as the 'good' strategies, and the Dishonor condition and the study Correct items were taken as bad strategies, with the former being due to self selection by the participant and the latter due to computer selection. This designation allowed us to subject the data to a 2 X 2 X 2 split plot analysis of variance, in which the factors were grade (3 or 5), strategy (good or bad) and source (Self selection or Computer selection).

There was a main effect of 'good strategy', such that overall, participants did better in the good than the poor strategy conditions ($F(1,21)=24.13$, $MSe=.03$, $p<.01$, $ES=.54$, $P=.99$). There was no effect of source ($F<1$, $p>.05$, $P=.15$), nor was there an interaction between source and goodness of strategy, ($F(1,21)=1.22$, $p>.05$, $P=.18$) or an overall effect of grade ($F<1$, $p>.05$, $P=.06$). However, as can be seen in Fig. 4 there was an interaction between grade and strategy, $F(1,21)=7.04$, $MSe=.03$, $p<.05$, $ES=.25$, $P=.72$. The fifth graders did considerably better with the Good strategies ($M=.82$, $SE=.05$), both in the Self condition and in the Computer condition, than they did with the Bad strategies ($M=.56$, $SE=.03$). The third graders, showed a smaller advantage than did the 5th graders with the Good strategies versus Bad strategies in the Computer condition and no advantage with the Good (honor choice) strategy, as compared to the Bad (dishonor choice) strategy, in the Self condition ($t<1$, *n.s.*).

Choices As might be inferred from the above performance data, the fifth graders tended to choose, appropriately, to study the items they did not already know (i.e., the incorrect items). 82 % of their choices were of items on which they had been incorrect on the initial test. This was significantly more than 50 %, $t(8)=10.00$, $p<.01$. Third graders, by contrast, tended to choose randomly. Their percentage of choices that were of the incorrect items was 60 %, and this was not different from 50 % $t(13)=1.40$, *n.s.*

Discussion

The situation presented in this experiment was the most straightforward choice that could enhance study. The correct choice, in this situation, is agreed upon by all theories: people should not study the items they know already, but rather the items that they do not know. When the computer enforced this choice, children in both grades showed enhanced recall.

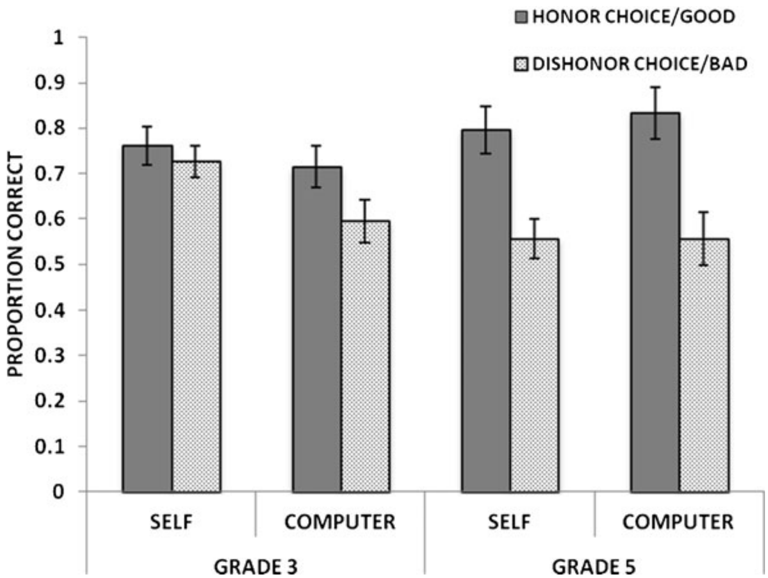

Fig. 4 Accuracy for good and bad strategies as a function of whether the strategy was self-selected or computer selected, in Experiment 5. Error bars indicate the standard errors of the mean

However, despite the good metaknowledge exhibited by both Grade 3 and Grade 5 children in the first four experiments, this experiment showed that only the Grade 5 children used that knowledge—in this most obvious of situations—to determine what to study. The younger children made random study choices, and failed to benefit from having their choices honored.

Experiment 6

In this final experiment we investigated a more subtle study-choice discrimination than the know/don't know choice of the previous experiment. We investigated study choices when *all* of the known items were eliminated. Choices were restricted to being among items that were not yet learned, but which had different degrees of learning. The Region of Proximal Learning model and the Discrepancy Reduction models make different predictions about what people should choose under these conditions. The Region of Proximal Learning model says that the optimal choice, in such a situation, is to first study the items that are closest to being learned, turning to the more difficult and less well learned items that would take more time and effort only after these nearly learned items have been studied. In contrast, the Discrepancy reduction model says that the person should choose to study the most difficult items—those furthest from being learned. Kornell and Metcalfe (2006) showed that adults choose to study the easiest-as-yet-unlearned items and benefit most from studying them. Our question here was whether the older children, in particular, might also choose to study the easiest-as-yet-unlearned items, and would benefit most from their selection for study (as is consistent with the Region of Proximal Learning model), or whether they would choose to study the most difficult items, as the Discrepancy Reduction model proposes, and as study choices in Experiment 4 and 5 suggest. We will also look at whether these choices are effective, by using the honor dishonor paradigm. Finally, we will investigate the consequences for learning when the computer selects the easiest as opposed to the most difficult unlearned items for study. Is performance enhanced when the easiest as yet unlearned items are studied, as the Region of Proximal Leaning model predicts, or is the reverse result obtained, as the Discrepancy Reduction model predicts?

Method

Participants The participants were 20 children from the Emily Dickinson School in New York City. There were 8 Grade 3 participants, and 12 Grade 5 participants.

Design The design was a 2 X 2 X 2 split plot factorial design similar to that of the previous experiment. One major change in the terminology in this experiment was that what is called the good strategy in the computer choice conditions was to study the easiest as yet unlearned items, that is, the items with the highest judgments of learning. This terminology is consistent with the predictions of the Region of Proximal Learning model. Had we taken JOLs before study choices in the previous experiment, the optimal strategy would have been to study the items with the lowest JOL. In this experiment, however, the optimal choice is to study the items with the highest JOL. Here we will call Honor choice, and study the *highest* JOL items selected by the computer the 'good' strategies, based on these theoretical considerations.

Procedure In the pretest the children were asked for the English translations of Spanish words, until 24 incorrect answers had been amassed. During the pretest the Spanish word was presented and the child was asked to make a JOL by using a slider as in the previous experiments. After making the judgment, the child was asked to provide the English translation, after which the next Spanish vocabulary word was presented.

Once 24 incorrects had been accumulated, the computer rank ordered these questions according to the children's JOLs on each, and then randomly assigned items from among the highest four JOL items to the four conditions (honor choice, dishonor choice, high JOL computer choice, Low JOL computer choice). Then it went on to the next highest ranked four items and randomly assigned them to condition, and so on until all 24 items had been assigned. This assignment procedure meant that the distributions of JOLs for the 6 items in each of the 4 conditions were close to being the same, and that the JOLs in each sextet spanned the entire range from highest to lowest JOLs. The four conditions were then assigned to order of presentation of choice, such that condition and order were counterbalanced across participants. The 6 items in each condition were randomized and arranged in a circle, and the children went through the 4 choice blocks, choosing the 3 items from each block that s/he wished to study in order to produce the best possible performance on the subsequent test on all 24 items. They were then given the study phase, which was just like that used in Experiment 5. After the study phase they were tested on all 24 items.

One small change from the previous experiment was made in the final testing procedure. Although we only counted the first attempt at recall in the data that are presented below, we repeatedly tested each child in an iterative procedure. Because the iterations were done after the data presented below were collected, they could not have changed the results. When a child made a mistake, rather than getting no feedback, s/he was given the correct answer, and required to produce it (as in the study phase). That item was then put back in the pool to be retested after all 24 items had been tested the first time. We did this recursively (with a stop rule of not continuing if an item was not recalled correctly after 4 complete tries) taking out items as they were correctly recalled—which meant the unlearned pool got smaller and smaller, and the spacing of repeated tries closer and closer. We did this because by the end of the experiment the children had learned—and got credit for in terms of a ping and a token presented onscreen—all or nearly all items. This resulted in a highly encouraging successful learning experience for the children (whereas if we had stopped after the first pass, they would have perceived themselves to have done poorly indeed). So, following Finn (2010), we wanted the children to end on a high note, and this procedure ensured that they did.

Results

Learning We, again, grouped conditions into 'good' and 'bad' strategy based on our previous research with adults and on the Region of Proximal Learning model. The good strategy included the honor choice condition and the select high JOL condition; the bad strategy included the dishonor choice condition and the select low JOL condition. There was a main effect of 'good strategy', such that overall participants did better in the good (i.e., the honor choice, or the select high JOL) than the poor strategy conditions, ($F(1,18)=6.82$, $MSe=.02$, $p<.05$, $ES=.28$, $P=.70$). However, this main effect was qualified by an interaction between source and strategy, $F(1,18)=4.29$, $MSe=.02$, $p=.05$, $ES=.19$, $P=.50$, such that in both Grade 3 and 5, the benefit of the 'good' strategy was entirely attributable to the computer choice, (self: $t<1$, *n.s.*, computer: $t(20)=3.07$, $p<.01$). The children did not spontaneously choose to study the items that would have helped most. Children in neither grade benefited from having their choices honored, though when the computer chose the items that the Region of Proximal Learning model indicates as being the best items for study, recall performance was enhanced. See Fig. 5.

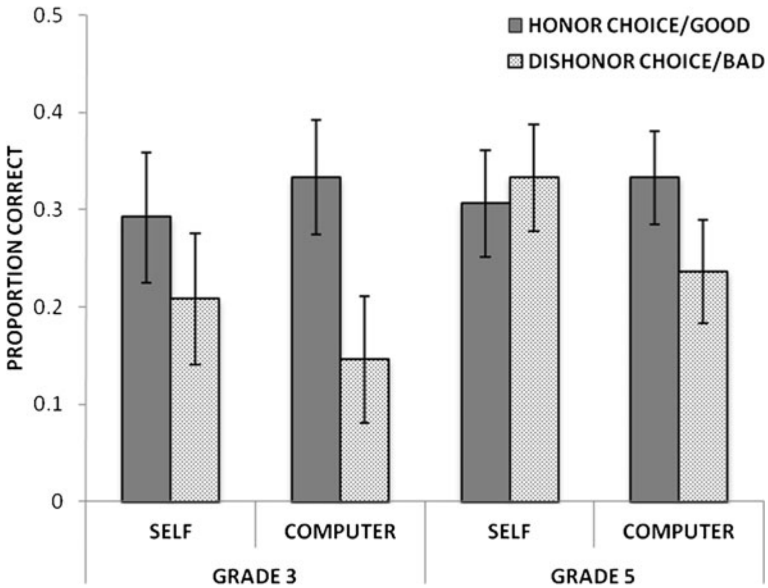

Fig. 5 Accuracy for good and bad strategies as a function of whether the strategy was self-selected or computer selected in Experiment 6. Error bars indicate the standard errors of the mean

Choices The mean JOL of the items the third graders chose was 0.44, *SE*=.05, whereas it was 0.45, *SE*=.05 for the items that they did not choose. This difference was not significant $t<1$, $p>.05$. The mean JOL for the items the fifth graders chose was 0.41, *SE*=.07. For the items they declined to choose it was 0.59, *SE*=.07. This difference was significant, $t(11)=2.34$, $p<.05$. However, notice that the Grade 5 children chose the low rather than the high JOL items, that is, their choices were in the direction that *hurt* rather than helped their recall (and were the opposite of those condoned by the Region of Proximal Learning model).

Discussion

In the present more difficult discrimination task, neither the third graders nor the fifth graders provided evidence of making an effective choice of items to study, unlike adults. This choice task, though, is considerably more difficult that the simple discrimination between known and not known in which the fifth graders, in the previous experiment, showed evidence of beginning to have an informed choice strategy in place. Indeed, though the fifth graders showed some evidence of making selective choices, they were the wrong choices, and did not help recall.

Conclusion

The first three experiments showed that Grade 3 and Grade 5 children had highly accurate metacognitive judgments when tested on a paradigm in which adults show highly accurate metacognitive judgments. The children did exhibit some overconfidence and wishful thinking, but both appeared to be more related to their level of performance than to their age. When the children performed well, overconfidence and wishful thinking were minimized. When they performed poorly, regardless of grade, they tended to be overconfident and to exhibit wishful thinking that they had already learned items that they did not yet know. We found no relation between the goodness of the metacognitions, as given by an individual's relative accuracy gamma, and the goodness of performance, as given by recall.

There may be several reasons for this, including the high and range-restricted levels of metacognitive accuracy, and the fact that in the first three experiments, even if the children knew that they did not yet know something, they were not given an opportunity to implement remedial action. We are not arguing that there is never a relation between good metacognition and good performance. We do suggest, in accord with Thiede et al. 2003, that for good metacognition to result in improved performance, people have to both have accurate metacognition, and also be in a position to use that knowledge to enact good study strategies. And, they need to know what the good strategies are. And they need to use them. Good metacognition alone does not automatically result in high performance levels.

The last three experiments showed that despite adequate metaknowledge, children at this age are in flux about using this knowledge to benefit study. The younger children exhibited a distinct implementation deficit. Even in the simplest case in which all they needed to do (in Experiment 5) was to choose to study the items they did not know as compared to the ones that they did know, the younger children chose randomly. The older children, by contrast, were able to choose appropriately in this situation, and their choices benefitted their recall. However, even the older children showed an implementation deficit when the discrimination task was made more difficult (in Experiment 6). In this case, recall was enhanced by study of the easiest as yet unlearned items, or by the items on which the children had expressed their highest judgments of learning. Studying these items benefited the children's learning as evidenced by the conditions in which the computer selected these items for study. In these computer-controlled conditions both the younger and the older children's recall benefitted as compared to studying the items with the highest JOLs. But neither the younger nor the older children, themselves, chose these items for restudy.

We do not, as yet, know why the younger children do not use their metacognitive knowledge, and why the older children begin to use it, but not always optimally. It is possible that though both groups exhibit good metacognitive knowledge, their metacognitive experiences might be different (see Efklides 2011 and Schwartz and Metcalfe 2011, for a review of this distinction). The younger and older children's *feelings* about their learning, which may be necessary to allow them to use their metacognition strategically to determine choices, may be different. Alternatively, the younger children may simply fail to remember or recruit their metacognition, when the time comes to make study choices, and may make those choices based on other considerations.

It is interesting to consider the strategies that the Grade 5 (but not the Grade 3) children were beginning to use—presumably the first study choice strategies that are spontaneously available to them—from the theoretical perspectives of the Region of Proximal Learning and the Discrepancy Reduction models. In the case, illustrated in Experiment 4, in which the study choices were among a wide variety of materials, including well-learned and unlearned items, the Region of Proximal Learning model says that people should decline study of already-learned items. Use of this strategy will result in a negative correlation between items chosen and JOLs. Such a negative correlation was shown by the Grade 5 children. The Discrepancy Reduction model says that people should study the most difficult items. This, however, also results in a negative correlation between items chosen and JOLs. In Experiment 5, in which half of the items were already-learned items and half of them were unlearned, the Region of Proximal Learning model says that people should decline to study the already-learned items. The Discrepancy Reduction model says they should selectively study the lowest JOL items. The result of both of these strategies is that people should choose study of the most difficult items which are the as-yet-unlearned items. This is what the Grade 5 children did. However, in Experiment 6, when the choice was among only unlearned items, the two models make opposite predictions. The Discrepancy Reduction Model still says that the most difficult items should be chosen. However, the Region of Proximal Learning model says that it is optimal to choose the *easiest* as-yet-unlearned items. (And indeed, when the computer chose these easiest items for the children to study their learning was enhanced, vouching for the optimality of the strategy). The choice pattern that the children, themselves, displayed, though, was that of the Discrepancy Reduction Model, not that of the Region of Proximal Learning model. The fact that these two models are explicit about the strategies that people might be using is of considerable interest in helping to isolate the details of the development of strategic metacognitively guided study choice. Our data indicate that the earliest strategy implemented is that of the Discrepancy Reduction model. Much of the time, but not always, this nascent strategy helps learning. However, as is illustrated in Experiment 6, sometimes it does not.

There are three limitations to these studies. The first limitation was that the sample sizes were small, especially in the first three experiments. It would, of course, have been desirable to have had a larger sample. However, even with the small sample sizes, the primary result—showing excellent relative accuracy metacognition both at the grade 3 and the grade 5 level across a variety of materials—replicated every time. The children's relative accuracy was at about the same level of that of adults and was near ceiling.

A second limitation was that some of the children were Hispanic whereas others were not, and in some of the experiments, we used Spanish/English materials. This could, potentially, have caused a bias. We do not think it did, for two reasons. First, the 'easy' items in the Spanish materials were cognates, such as familia-family or tomate-tomato, which were easy regardless of whether the individual translated from Spanish to English or the reverse. Intermediate items included pairs such as viento-wind, which might have been known in one language but not the other, for both groups of children. The most difficult items were pairs like choquezuela-kneecap, many of which were unknown even to adult Spanish speakers, and would have been difficult for all children. Thus, it was not clear that either ethnic group was at an advantage, given these materials. Second, differences in the amount of prior knowledge is the norm in studies of semantic metacognition and is present even in the experiments with general knowledge or vocabulary items. The RPL framework specifically takes this prior knowledge issue into account, proposing that choices need to be tailored to the individual's own knowledge. We did this, especially in Experiment 6, by eliminating those items that the participants got right on a prior test. This would have put the more and less knowledgeable children on an even footing. A third limitation is that we only looked at study choices that were based on relative metacognitive accuracy, rather than absolute accuracy. Because of the constraints of our design, they had to choose exactly half of the items for restudy, and our question was which items did they choose, rather than how many, or, indeed, whether they would choose to restudy any of them at all if they had not been constrained. Thus, we could not pick up age-related differences in the overall choices to study or to decline study entirely. The strategies evaluated here related only to whether the children chose to study the items that would benefit from study rather than those that would not, when they had to choose among items. Finn (2008) and Metcalfe and Finn (2008) have shown that adults' absolute level of confidence influences whether or not they choose to restudy at all, such that the individual is more likely to choose to restudy when they are under confident rather than overconfident. While we could not look at such freely determined choices given the constraints and objectives of the present design, the issue of whether there are developmental trends in study choices based on absolute confidence certainly deserves further scrutiny.

Overall, these studies reveal a particular profile of metacognitive/control development. By Grade 3, the children's metacognitions, evaluated in terms of their relative accuracy, are fully developed, at least when they are tested using a methodology in which adults show highly accurate metacognitions. Children may still be developing more subtle heuristics in other, less well-defined, paradigms at this age. But even in maturity, those other paradigms—such as the immediate JOL paradigm—are not very predictive of later performance. Although there may still be metacognitive nuances and heuristics that children at this age have yet to learn, the metacognitive fundamentals needed for most classroom situations appear to be in place—they know which items are known better than others. Teachers need to be aware that children can make these judgments of which items they know and do not yet know accurately, if they are queried using the delayed JOL paradigm. They could instruct their students to use this technique (which has other study benefits, see Kimball and Metcalfe 2003).

The effective implementation of study strategies based even on this metacognitive knowledge, though, cannot be assumed at this critical stage of middle childhood. A change in implementation is seen between Grade 3, in which there is a pronounced implementation deficit, and choices appear to be essentially random, and Grade 5, in which the children use a strategy of choosing the lowest JOL items, but use it regardless of whether it is optimal for learning or not. An instructor or parent would do well to check how and if their students in this age range are implementing their metaknowledge appropriately. Their learning might be well enhanced with strategic selection of what they should study, and their own choices on this may not be reliable. Once effective study choice is being implemented reliably by the child, however, it would seem reasonable to accord the child more autonomous control over his or her own learning process with some assurance that such freedom will bode well for their learning.

Acknowledgements We thank Demian Vanderputten, Lisa Son, Umrao Sethi, Mike Serra, Camille Williams, Ljubica Chatman, Matthew Greene and the children, principal, teachers and parents of P.S. 75, The Emily Dickenson School, for their help. The research reported here was supported by James S. McDonnell Foundation Grant 220020166 and the Institute of Education Sciences, U.S. Department of Education, through Grant R305H060161 to Janet Metcalfe and Lisa Son at Columbia University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: when agendas override item-based monitoring. *Journal of Experimental Psychology. General, 138*, 432–447. doi:10.1037/a0015928.

Arnold, L. S. (2009). Enhancing student academic regulatory processes: A study of metacognition, strategic enactment and achievement. In Paul R. (Ed.), Pintrich Award Invited Address, American Psychological Association Convention, Toronto, Canada

Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology, 96*, 124–129. doi:10.1037/h0033475.

Begg, I. M., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: how useful is self-knowledge about memory? *European Journal of Cognitive Psychology, 4*, 195–218.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology. General, 127*, 55–68.

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 73–94). New York: Psychology Press.

Berlyne, D. E. (1978). Curiosity and learning. *Motivation and Emotion, 2*, 97–175.

Bisanz, G. L., Vesonder, G. T., & Voss, J. F. (1978). Knowledge of one's own responding and the relation of such knowledge to learning: a developmental study. *Journal of Experimental Child Psychology, 25*, 116–128.

Bjorklund, D. F., & Douglas, R. N. (1997). The development of memory strategies. In N. Cowan (Ed.), *The development of memory in childhood* (pp. 201–246). Hove East Sussex: Psychology Press.

Brown, A. L., Bransford, J., Ferrara, R., & Campione, J. (1983). Learning, remembering, and understanding. In P. H. Musen (Ed.), *Handbook of child psychology* (Vol. III, pp. 77–166). New York: Wiley.

Butterfield, E., Nelson, T. O., & Peck, J. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology, 24*, 654–663.

Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: a critical examination. *Child Development, 53*, 11–28.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12*, 50–71.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research. *Review of Educational Research, 76*, 1–62.

Cultice, J. L., Somerville, S. C., & Wellman, H. M. (1983). Preschooler's memory monitoring: FOK judgments. *Child Development, 54*, 1480–1486.

de Bruin, A., Thiede, K. W., Camp, G., & Redford, J. (2011). Generating keywords improves metacomprehension and self-regulation in elementary middle school children. *Journal of Experimental Child Psychology, 109*, 294–310.

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: differential and sufficient aspects. *Journal of Experimental Child Psychology, 47*, 274–296.

Dunlosky, J & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. Ross (Ed), *Psychology of Learning and Motivation, 54*, 103–140.

Dunlosky, J., Ariel, R., & Thiede, K. W. (2011). Agenda-based regulation of study-time allocation. In P. A. Higham & J. P. Leboe (Eds.), *Constructions of remembering and metacognition: Essays in Honour of Bruce Whittlesea* (pp. 182–198). Basingstoke: Palgrave MacMillan.

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multi-trial learning. *Journal of Gerontology: Psychological Sciences, 52*, 178–186.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–276). Mahwah: Erlbaum.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33*, 545–565.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect people's self-paced study and yield "labor-and-gain" effects. *Acta Psychologica, 98*, 37–56.

Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory and Cognition, 32*, 779–788.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist, 46*, 6-25.

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory and Cognition, 36*, 815–821.

Finn, B. (2010). Ending on a high note: adding a better end to difficult study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1548–1553.

Flavell, J. H., Frederichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology, 1*, 324–340.

Hacker, D. J., Bol, L., & Keener, M. C. (2010). Metacognition in education: A focus on calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of memory and metacognition*. Mahwah: Lawrence Erlbaum Associates.

Hebb, D. O. (1949). *Organization of behavior*. New York: Wiley.

Hofferth, S. L., & Sandberg, J. F. (2000). How American children spend their time. *Journal of Marriage and Family, 63*, 295–308.

Kao, Y.-C., Davis, E. S., & Gabrieli, J. D. E. (2005). Neural correlates of actual and predicted memory formation. *Nature Neuroscience, 8*, 1776–1783.

Kelly, M., Scholnick, E. K., Travers, S. H., & Johnson, J. W. (1976). Relations among memory, memory appraisal, and memory strategies. *Child Development, 47*, 648–659.

Kelemen, W. L., & Weaver, C. A. (1997). Enhanced metamemory at delays: why do judgments of learning improve over time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1394–1409.

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory and Cognition, 31*, 918–929.

Koriat, A. (2002). Metacognition research: An interim report. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 261–286). Cambridge: Cambridge University Press.

Knouse, L. E., Anastopoulos, A. D., & Dunlosky, J. (2013). Isolating metamemory deficits in the self-regulated learning of adults with AD/HD. *Journal of Attention Disorders.*

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology. General, 131*(2), 147–162.

Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy: Insights from the processes underlying judgments of learning in children. In P. Chambres, M. Izaute, & P.-J. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 1–17). New York: Kluwer.

Kornell, N. (2005). The effectiveness of study and the region of proximal learning. *International Dissertation Abstracts* (UMI No. 3174827).

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 609–622.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

Lockl, K., & Schneider, W. (2002). Developmental trends in children's feeling-of-knowing judgments. *International Journal of Behavioral Development, 26*, 326–333.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 663–679.

Markmann, E. M. (1977). Realizing that you don't understand: a preliminary investigation. *Child Development, 48*, 986–992.

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: why is study time sometimes not effective? *Journal of Experimental Psychology. General, 122*, 47–60.

Masur, E. F., McIntyre, C. W., & Flavell, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrial free recall task. *Journal of Experimental Child Psychology, 15*, 237–246.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology. General, 131*, 349–363.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*, 159–163.

Metcalfe, J. (2010). Desirable difficulties and studying in the Region of Proximal Learning. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork*. London: Psychology Press.
Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin Review, 15*, 174–179.
Metcalfe, J., & Jacobs, W. J. (2010). People's study time allocation and its relation to animal foraging. *Behavioral Processes, 83*, 213–221.
Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology. General, 132*, 530–542.
Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*, 463–477.
Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory and Cognition, 37*, 1077–1087.
Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based program to enhance study efficacy in a high and low-risk setting. *European Journal of Cognitive Psychology, 19*, 743–768.
Monroe, E. K., & Lange, G. (1977). The accuracy with which children judge the composition of their free recall. *Child Development, 48*, 381–387.
Nelson, T. O. (1984). A comparison of current measures of feeling-of-knowing accuracy. *Psychological Bulletin, 100*, 128–132.
Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science, 2*, 267–270.
Nelson, T. O., Dunlosky, J., Graf, P. A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5*, 207–213.
Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). New York: Academic.
Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: monitoring learning from text. *Educational Psychologist, 25*, 19–33.
Pressley, M., Levin, J. R., & Ghatala, E. S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning and Verbal Behavior, 23*, 270–288.
Pressley, M., Levin, J. R., Ghatala, E. S., & Ahmad, M. (1987). Test monitoring in young grade school children. *Journal of Experimental Child Psychology, 43*, 96–111.
Reder, L. M. (1988). Strategic control of retrieval strategies. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 227–259). San Diego: Academic.
Richards, R. M., & Nelson, T. O. (2004). Effect of the difficulty of prior items on the magnitude of judgments of learning for subsequent items. *The American Journal of Psychology, 117*, 81–91.
Roebers, C. M., von der Linden, N., Schneider, W., & Howie, P. (2007). Children's metamemorial judgments in an event recall task. *Journal of Experimental Child Psychology, 97*, 117–137.
Scheck, P., & Nelson, T. O. (2005). Lack of pervasiveness of the underconfidence-with-practice effect: boundary conditions and an explanation via anchoring. *Journal of Experimental Psychology. General, 134*, 124–128.
Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (2nd ed.). Mahwah: Erlbaum.
Schneider, W., Vise, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: evidence from a judgment-of-learning task. *Cognitive Development, 15*, 115–134.
Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*, 351–371.
Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition, 39*, 737–749.
Son, L. K. (2005). Metacognitive control: children's short-term versus long-term study strategies. *The Journal of General Psychology, 132*, 347–363.
Son, L. K. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 255–262.
Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204–221.
Son, L. K. (2005). Metacognitive control: Children's short-term versus long-term study strategies. *Journal of General Psychology, 132*, 347–363.
Son, L. K. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 255–262.
Son, L. K. & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204–221.
Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin Review, 6*, 662–667.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66–75.
Vygotsky, L. S. (1987). The collected works of L. S. Vygotsky, Volume 1. Problems of general psychology including the volume thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), New York: Plenum Press.
Weaver, C. A., & Keleman, W. L. (1997). Judgments of learning at delays: shifts in response patterns or increased metamemory accuracy? *Psychological Science, 8*, 318–321.
Wellman, H. M. (1977). Tip of the tongue and feeling of knowing experiences: a developmental study of memory monitoring. *Child Development, 48*, 13–21.
Wellman, H. M. (1978). Knowledge of the interaction of memory variables: a developmental study of metamemory. *Developmental Psychology, 14*, 24–29.
Wellman, H. M., Collins, J., & Glieberman, J. (1981). Understanding the combination of memory variables: developing conceptions of memory limitations. *Child Development, 52*, 1313–1317.
Woodworth, R. S. (1921). *Psychology: A study of mental life*. NY: Henry Holt and Company.
Worden, P. E., & Sladewski-Awig, A. (1982). Children's awareness of memorability. *Journal of Educational Psychology, 74*, 341–350.

Author Note

Janet Metcalfe, Department of Psychology, Columbia University; Bridgid Finn, Department of Psychology, Washington University in St. Louis. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through CASL grant #R305H030175, and by Grant 220020166 from the James S. McDonnell Foundation. The authors are entirely responsible for the content of this article. We also thank Lisa Son, David Miele, Barbie Huelser, and the anonymous reviewers. Correspondence should be addressed to Janet Metcalfe, Department of Psychology, Columbia University, New York, NY 10027. Email: jm348@columbia.edu