Metacognition Learning
https://doi.org/10.1007/s11409-018-9180-y

Study techniques differentially influence the delayed judgment-of-learning accuracy of adolescent children and college-aged adults

Gregory I. Hughes[1] **· Holly A. Taylor**[1] **· Ayanna K. Thomas**[1]

Received: 23 February 2017 / Accepted: 2 January 2018

Abstract The ways in which adult learners study information influences their judgment-of-learning (JOL) accuracy (e.g., Koriat et al. *Journal of Experimental Psychology: General, 131*(2), 147-162, 2002). The present study extends this investigation to adolescent children to determine whether developing learners' metacognitive monitoring is similarly influenced by different study techniques. In two experiments, we examined JOL accuracy in adolescent children (ages 11 to 12) and college-aged adults. Across both experiments, we employed a standard delayed-JOL paradigm in which three groups of participants, differing in type of study technique, encoded weakly-related word pairs. One group studied the word pairs twice (study practice). A second group studied with the instruction to generate a word that linked the two members of the word pair together (elaborative encoding). The final group studied word pairs and then took an immediate cued-recall test with feedback (retrieval practice). In children and adults, retrieval practice led to better JOL accuracy as compared to study practice. Children differed from adults in how elaborative encoding influenced JOL accuracy. For adults, elaborative encoding resulted in better JOL accuracy than study practice; however, for children, JOL accuracy did not differ between the two groups. Our results suggest that encoding processes influence delayed-JOL accuracy in both age-groups.

Keywords Metamemory · Children · Judgments of learning · Monitoring · Retrieval practice

Students entering middle school are increasingly tasked with managing their learning outside of the classroom. Without teacher supervision, they must rely on metacognitive processes to guide crucial study decisions, such as what, how, and how long to study (Nelson and Narens 1994). *Metacognition* refers to the dynamic process in which individuals introspect on learning

✉ Gregory I. Hughes
gregory.hughes@tufts.edu

1 Department of Psychology, Tufts University, 490 Boston Ave, Medford, MA 02155, USA

and memory processes (*monitoring*) and then use that information to make decisions about how to regulate those processes (*control*; Nelson and Narens 1990). Research has consistently demonstrated that monitoring influences control (for a review, see Metcalfe 2009), and that monitoring accuracy influences the efficacy of self-regulated, or controlled study (Thiede 1999; Thiede et al. 2003). Unfortunately, research has also consistently demonstrated that monitoring is often inaccurate, and such inaccuracies have downstream consequences for control (Karpicke 2009; Thomas and McDaniel 2007). In the present experiment, we explored whether different study techniques influenced judgment-of-learning (JOL) accuracy in adolescent children (ages 11 to 12) and college-aged adults.

A JOL is a prediction about the likelihood of later remembering recently studied material. In a typical JOL experiment, learners study word pairs and then, on a numeric scale, estimate the likelihood that they will remember the target (second member of the pair) when later prompted with the cue (first member on the pair; for a review see Rhodes 2016). Finally, learners take a cued-recall or recognition test, which allows researchers to assess how well JOLs predict subsequent memory performance.

JOL accuracy has traditionally been measured either by (a) calculating Goodman and Kruskal Gamma (γ) correlations between JOLs and later memory performance on an item-by-item basis (*relative accuracy*), or (b) by comparing mean JOLs to mean memory performance (*absolute accuracy*; for a review see Rhodes 2016). Relative accuracy is perfect when the JOLs for items that are later remembered are always higher than the JOLs for items that are later forgotten. Absolute accuracy is perfect when there is no difference between mean JOLs and mean performance. In the present study, we focused on the factors that influence relative accuracy (hereafter referred to simply as JOL accuracy) because it is critical to the effective allocation of study time across items.

Multiple factors influence JOL accuracy. Delaying JOLs from the initial study session is perhaps the most effective and reliable method of enhancing JOL accuracy. Using a paired-associate paradigm, Nelson and Dunlosky (1991) were the first to observe the *delayed-JOL effect*, in which JOLs made at a short delay after studying items were substantially more accurate than JOLs made immediately after studying. The delayed-JOL effect is highly robust and has been widely replicated with a variety of materials and designs (Rhodes and Tauber 2011). In a meta-analysis, Rhodes and Tauber (2011) found that the mean weighted accuracy of delayed JOLs, as measured by item-by-item γ correlations between predictions and memory performance, was .77, whereas it was only .42 for immediate JOLs. Delay not only improves JOL accuracy in adults, but also in school-aged children (Koriat and Shitzer-Reichert 2002; Schneider et al. 2000; van Loon et al. 2013; but see Roebers et al. 2007).

Another means of influencing JOL accuracy involves targeting how learners initially study material. A broad body of research conducted with adults has shown that immediate JOL accuracy increases with the number of times learners review material, such as through repeated study (Lovelace 1984; Mazzoni et al. 1990; but see Meeter and Nelson 2003; Koriat 1997) or study-test practice (i.e. *retrieval practice*; Koriat 1997; Koriat et al. 2002; Koriat et al. 2006; Koriat and Shitzer-Reichert 2002; Leonesio and Nelson 1990; Jang et al. 2012). In these studies, JOLs are made at each study or study-test phase and are compared to a subsequent final test to assess accuracy. Notably, some study techniques result in higher JOL accuracy than others. For example, engaging in retrieval practice during learning results in higher JOL accuracy compared to simply studying material (Lovelace 1984; King et al. 1980; Shaughnessy and Zechmeister 1992; but see Meeter and Nelson 2003).

The goal of the present study was to investigate whether different study techniques influenced the accuracy of delayed JOLs in adolescent children and college-aged adults. Such an investigation has both theoretical and practical significance. From a practical perspective, designing metacognitive interventions for student populations should capitalize on the benefit afforded by delay because it is the most robust, reliable, and efficient means of enhancing the accuracy of JOLs. Consider that Koriat et al. (2006) found that it takes multiple study-test trials for the immediate-JOL accuracy to match delayed-JOL accuracy following only a single study-test trial. If certain study techniques can enhance the accuracy of JOLs beyond the potent benefits of delay, then this can yield highly efficient interventions. From a theoretical perspective, this investigation will provide important information about the factors that influence the accuracy of delayed JOLs in children and adults. Differences in the influence of study techniques on the JOL accuracy between the age groups would provide insight into the development of processes underlying metacognitive monitoring predictions.

Compared to immediate JOLs, different kinds of study techniques have not been shown to influence the delayed-JOL accuracy of adults. However, the few studies that exist do not conclusively negate the possibility. Although researchers have observed no differences in delayed-JOL accuracy when comparing interactive imagery to study practice (Begg et al. 1989), interactive imagery to separate imagery (Dunlosky and Nelson 1994), and a single study trial to two spaced study trials (Dunlosky and Nelson 1994), these null results are not surprising because comparing these study techniques has not reliably led to differences in immediate-JOL accuracy. However, other studies have compared study techniques that lead to differences in immediate-JOL accuracy and still did not find differences in delayed-JOL accuracy (Koriat et al. 2006; Jang et al. 2012). For example, Koriat et al. (2006) found that repeated retrieval practice did not result in increasingly higher delayed-JOL accuracy. Likewise, Jang et al. (2012) found no benefit of retrieval practice over study practice on delayed-JOL accuracy. However, the authors of both studies noted that delayed-JOL accuracy was at ceiling after the first study or retrieval practice trial, leaving no room for additional trials to benefit JOL accuracy or for differences to emerge between the two study techniques. In stark contrast to these studies, Peynircioğlu et al. (2014) found that delayed JOLs for short piano pieces were more accurate if those pieces were initially learned visually, or visually and aurally, compared to only aurally. Admittedly, there are many important differences between the previous studies that did not find an effect of study technique on JOL accuracy and the one presented by Peynircioğlu et al. (2014). Stimuli, type of study, and general modality differences make it challenging to directly compare this study with those that employ standard verbal-learning methods. However, these results suggest that how learners encode material may influence their delayed JOLs for that material.

Few studies have examined the influence of study techniques on immediate- or delayed-JOL accuracy in children. Regarding immediate-JOL accuracy, the handful of studies that exist present mixed findings. For example, although Koriat and Shitzer-Reichert (2002) found that repeated retrieval practice benefited the immediate-JOL accuracy of 7- and 10-year-olds, Finn and Metcalfe (2014) did not replicate this finding in 9- and 10-year-olds with a highly similar design. Tsalas et al. (2015) found that repeated study benefited the JOL accuracy of 7-year-olds, but not that 10-year-olds or adults, likely because 10-year-olds' and adults' JOL accuracy after the first trial was near ceiling.

Regarding the influence of study techniques on delayed-JOL accuracy, the two studies conducted with children provide predominantly null findings. However, these studies, like experiments conducted with adults, do not firmly contradict the possibility. Although van

Loon et al. (2013) found that a sentence generation technique did not lead to higher delayed-JOL accuracy than simple study, these techniques have not been shown to lead to differences in immediate-JOL accuracy in adults, and thus these results are not surprising. Von der Linden et al. (2015) found that using interactive imagery with paired associates led to higher delayed-JOL accuracy compared to study practice in 8-year-olds, but not in 12-year-olds, college-aged adults, or older adults. It is thus possible that interactive imagery does not benefit immediate-JOL accuracy in adolescent children and adults, but does support the JOL accuracy of younger children.

It is unclear whether study techniques will influence the accuracy of delayed JOLs in 11- to 12-year-old children and college-aged adults in the same way. There is evidence to suggest that there will be no developmental pattern because, at least in paired-associate paradigms, processes underlying JOL accuracy are well-developed, even comparable to adults, by roughly the ages of 7 to 10. Several studies have observed no developmental differences in JOL accuracy of children within this age-range (see, Finn and Metcalfe 2014; Metcalfe and Finn 2013; Roebers et al. 2007; Schneider et al. 2000; Tsalas et al. 2015; van Loon et al. 2013; von der Linden et al. 2015), or between children in this age-range and adults (Roebers et al. 2007; Tsalas et al. 2015). Further, similar factors influence the JOL accuracy of children in this age-range and adults alike, such as delay from encoding (Koriat and Shitzer-Reichert 2002; Schneider et al. 2000; van Loon et al. 2013; but see Roebers et al. 2007) and item difficulty (Koriat et al. 2009; Koriat and Shitzer-Reichert 2002; Tsalas et al. 2015).

However, there is evidence to suggest that JOL accuracy continues to develop through this age-range and into adolescence. For example, using a paired-associate task, Hoffmann-Biencourt and colleagues found that the JOL accuracy of 11- to 14-year-olds was higher than 7- to 9-year-olds after encoding material through multiple cycles of retrieval practice (Hoffmann-Biencourt et al. 2010). This pattern was also observed by Koriat and Shitzer-Reichert (2002) with 10- and 7-year-olds in a retrieval-practice paradigm. In addition, Finn and Metcalfe (2014) found that although 8- to 10-year-olds modulated their JOLs based on their experience with retrieval practice, they did not do so as effectively as what has been observed in adults. As a result, the accuracy of their JOLs did not increase across trials as has been consistently demonstrated with adults.

Whether a study suggests developmental changes in JOL accuracy through the age-range of 7 to 10 appears to depend on the complexity, or difficulty, of the encoding task. Studies that suggest no development through this age-range used passive encoding tasks, such as viewing a video (Roebers et al. 2007) or simply reading materials (Metcalfe and Finn 2013; Schneider et al. 2000; Tsalas et al. 2015, von der Linden et al. 2015; but see the interactive-imagery condition in von der Linden et al. 2015). In contrast, the studies that do suggest developmental trends through this age-range used retrieval practice (Finn and Metcalfe 2014; Hoffmann-Biencourt et al. 2010; Koriat and Shitzer-Reichert 2002), which is comparatively a more demanding and difficult encoding task. This pattern of results suggests that when predicting memory performance, the ability to use the experiential cues engendered from complex or difficult encoding procedures increases with age.

Experiment 1

In the first experiment, we explored the delayed-JOL accuracy of 11- to 12-year-old students in a paired-associate task. We compared delayed-JOL accuracy as a function of whether learners encoded information through one of three study techniques: study practice, retrieval practice with feedback, or an elaborative encoding technique. Study practice involved simply reading word pairs twice, elaborative encoding required that learners generate *mediator* words that thematically related the cues and targets of paired associates, and retrieval practice entailed taking practice tests on previously studied information and then receiving feedback in the form of the correct answer after every trial. Contrary to most JOL studies, we had learners predict associative-recognition rather than cued-recall. We did so because predictions of recognition are more difficult and typically less accurate than predictions of cued-recall (Thiede and Dunlosky 1994; Weaver and Kelemen 2003), thus minimizing the chances of ceiling levels of JOL accuracy that might mask the influence of study techniques (cf. Koriat et al. 2006; Jang et al. 2012).

We hypothesized that retrieval practice would lead to higher delayed-JOL accuracy than study practice because (a) retrieval practice consistently leads to higher immediate JOL accuracy in adults (Koriat et al. 2006), and (b) learners in this age-range have been shown to use the feedback from tests to modulate their monitoring predictions. However, we made no predictions about the influence of elaborative encoding on JOL accuracy because similar elaborative techniques, such as interactive imagery techniques, have not influenced immediate JOL accuracy in adults relative to study practice (Begg et al. 1989; Dunlosky and Nelson 1994).

Method

Participants

Eighty-five sixth grade students (ages 11 to 12) from a United States school in Hanover, New Hampshire, volunteered to participate in the experiment. All participants identified as Caucasian. A total 45 girls and 40 boys participated in this study. We obtained consent from both the students and their parents or guardians. We randomly assigned 26 participants to the study practice group, 28 to the elaborative encoding group, and 31 to the retrieval practice group.

Materials

Study material consisted of twenty weakly-related, cue-target word pairs. The mean forward associative strength of the cue to target was 3% ($SD = 1.9\%$) per the University of South Florida Free Association norms (Nelson et al. 1998). These norms were also used to acquire three weakly-related words used as foils for each word pair on the final associative recognition test. The mean of forward associative strength from cue to foil was 6% ($SD = 4.8\%$). We used weakly-related word pairs because the influence of retrieval practice, relative to study practice, on memory performance is generally greater when forward associative strength is low (Carpenter 2009). We chose not to use completely unrelated word pairs because it might render the elaborative encoding task too difficult.

Design

We used a between-subjects design consisting of three study technique groups: study practice, elaborative encoding, and retrieval practice.

Procedure

Participants were run in groups of three to six on laptop computers programmed with E-Prime software (Version 2.1; Schneider et al. 2002). The procedure consisted of five phases: initial study, strategic learning, retention interval, judgment-of-learning (JOL), and final test. Before the experimental session began, participants engaged in a practice session, during which they studied four word-pairs, then made JOLs, and finally took a four-alternative forced-choice test, in which they were presented with each cue word, and were required to select the corresponding target word on a list of four choices. The purpose of the practice phase was to familiarize participants with the instructions and the tasks. Participants were told that the tasks in the practice phase and the experiment would be identical.

Phase 1 In Phase 1, which was identical across all three groups, participants studied 16 word-pairs. Participants first received written and oral instructions informing them that they would be see pairs of words one at a time for a short duration and that they should try to remember these words for a later memory test. Word pairs appeared one at a time for a 1000 ms each. Presentation order was randomized across participants.

Phase 2 Phase 2 immediately followed Phase 1, and differed across the three groups. For each group, participants received written and oral instructions. Presentation rates for each group were determined through pilot testing and were selected to avoid ceiling performance. As in Phase 1, Phase 2 trials were randomized across participants.

Study practice Participants in the study practice group read identical instructions as those presented in Phase 1. They then saw the same 16 word-pairs, one at a time, for 1000 ms.

Elaborative encoding Participants in the elaborative encoding group were told that they would be studying the word pairs again, but this time would be asked to type in a new word for each pair that related to both the cue and target word. Participants in the elaborative encoding group were presented with the 16 word-pairs one at a time, for 500 ms each. After each word pair, participants were asked to supply a word (the mediator) that thematically related the cue and the target (e.g., *Space* for the pair *Moon – Galaxy*). There was no time limit for responding. On average, it took participants 7229 ms to enter a mediator ($SD = 2096$ ms). Participants left 5.1% of responses blank.

Retrieval practice Participants in the retrieval practice group were told that they would be taking a test on the words they studied in Phase 1, in which they would see the cue and provide the target. Participants were encouraged to guess if they were not confident in a retrieved answer. The cue remained on the screen until an answer was provided and there was no time limit to respond. On average, it took participants 5809 ms to enter a response ($SD = 1841$ ms). Mean accuracy during retrieval practice was 47% ($SD = 19\%$) on the cued-recall test and participants left only 1.8% of responses blank. After each trial, participants saw the complete word pair for 500 ms as feedback. Thus, the mean time on a given retrieval practice trial was approximately 6300 ms.

Phases 3 and 4 Phase 3 immediately followed Phase 2, and consisted of a 5-min retention interval in which participants performed a non-verbal drawing task. Participants received printed images of snowflakes and were asked to hand draw the snowflakes on blank paper. Immediately after the retention interval, Phase 4 began. Participants were told that they would see the cue from each of the studied pairs and would be instructed to estimate their likelihood of recognizing the corresponding target word on a multiple-choice test that would occur in about 5 min. Each JOL was prompted by the cue and the instruction to predict recognition on the final test. They provided their JOLs on a scale of 0 (*will not remember*) to 10 (*will definitely remember*). Participants were told that a rating of 5 indicates moderate levels of confidence and were encouraged to use the entire range of the scale. Judgments-of-learning were self-paced. Average time to enter a JOL was 3493 ms ($SD = 4384$ ms) in the study practice group, 3058 ms ($SD = 2102$ ms) in the elaborative encoding group, and 3473 ms ($SD = 3084$) in the retrieval practice group.

Phase 5 Phase 5 consisted of a four-alternative forced-choice test. Cues from the 16 word-pairs were randomly presented one at a time. Participants were asked to select the corresponding target word from a list of three other foils by typing in a numeral, ranging from 1 to 4, which corresponded with potential answers. There was no time limit to respond. Average time to enter a response on each test item was 5156 ms ($SD = 2398$ ms) in the study practice group, 4248 ms ($SD = 2345$ ms) in the elaborative encoding group, and 3744 ms ($SD = 1923$) in the retrieval practice group.

Results

For all subsequent statistical tests, we used an alpha level of .05 to determine significance. Post-hoc tests were adjusted with a Bonferroni correction.

Final test performance and mean JOL magnitude

A one-way between-subjects analysis of variance (ANOVA) on mean final test performance revealed a main effect of study technique, $F\ (2, 82) = 12.65$, $p < .001$, $\eta_p^2 = 0.27$. As shown in Table 1, retrieval practice ($M = .84$) led to higher performance than study practice ($M = .61$), $t\ (57) = 5.02$, $p < .001$, $d = 1.64$, as did elaborative encoding ($M = .75$), $t\ (54) = 2.98$, $p = .026$, $d = 0.72$. The difference in final test performance between retrieval practice and elaborative encoding was not significant ($p > .05$).

A one-way between-subjects ANOVA on mean JOL magnitude revealed no main effect of group, $F\ (2, 82) = .84$, $p = .436$, $\eta_p^2 = 0.24$.

Table 1 Experiment 1. Results of children. Mean JOLs (for all items and as a function of memory success on the final test), Final Test Performance, and JOL Accuracy (mean γ correlation coefficient)

Study Technique Group	JOL (All)	JOL (Incorrect)	JOL (Correct)	Final Test	JOL Accuracy
Study Practice	5.46 (*1.89*)	4.94 (*2.13*)	5.76 (*1.94*)	.61 (*.16*)	.16 (*.47*)
Elaborative Encoding	5.84 (*1.17*)	5.70 (*1.78*)	5.94 (*1.29*)	.75 (.23)	.06 (*.50)*
Retrieval Practice	6.00 (*1.70*)	4.32 (*2.29*)	6.16 (*1.67*)	.84 (*.12*)	.37 (*.46)*

Standard deviations given in parentheses

Judgment of learning accuracy

To measure JOL accuracy, we calculated intra-individual γ correlations between each participant's JOLs and test accuracy on an item-by-item basis. These correlations range from −1.0 to 1.0, with higher values indicating better JOL accuracy than lower values. Such correlations were incomputable for six participants from the retrieval practice group, four from the elaborative encoding group, and one from the study practice group due to lack of variance in the JOLs or four-alternative forced-choice test performance. Thus, these 11 participants could not be included in the analysis.

A one-way between-subjects ANOVA on JOL accuracy revealed a main effect of group, $F(2, 71) = 3.13$, $p = .05$, $\eta_p^2 = 0.08$. As shown in Fig. 1, retrieval practice ($M = .37$) resulted in better JOL accuracy than elaborative encoding ($M = .06$), $t(49) = 2.58$, $p = .05$, $d = 0.67$, but not study practice ($M = .16$), $t(48) = 1.74$, $p = .31$, $d = 0.49$. The difference between elaborative encoding and study practice was not significant ($p > .05$).

We conducted three one-sample *t*-tests to assess whether mean γ values for each group were statistically greater than 0, which would indicate above-chance JOL accuracy. The mean γ correlation coefficient associated with the retrieval practice group ($M = .37$) statistically differed from 0, $t(24) = 4.10$, $p < .001$, $d = 0.82$. Neither the mean γ correlation coefficients of the study practice group ($M = .16$), $t(24) = 2.0$, $p = .06$, $d = 0.40$, nor the elaborative encoding group ($M = .06$), $t(23) = .55$, $p = .59$, $d = 0.11$, statistically differed from 0.

We examined the possibility that elaborative encoding resulted in low JOL accuracy because participants generated mediators during study that subsequently appeared as foils on the four-alternative forced-choice test. To test this possibility, we removed all cases in which elaborative encoding participants generated a foil as a mediator during study and subsequently selected that mediator as the answer on the four-alternative forced-choice-test. We then recomputed γ correlations for the elaborative encoding group, which resulted in a slightly lower correlation coefficient ($M = .01$).

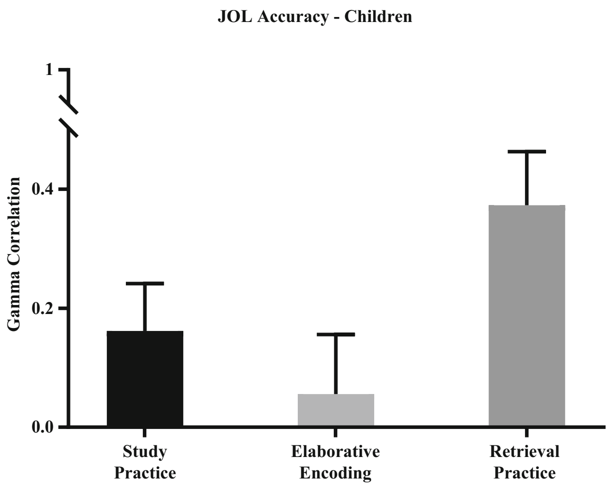

Fig. 1 Delayed judgment-of-learning accuracy of children, as measured by gamma correlation coefficients between JOLs and final test performance on an item-by-item basis. Error bars represent standard error of the mean

Experiment 1 discussion

Consistent with our predictions, we found that study techniques differentially influenced the delayed-JOL accuracy of children. Retrieval practice led to above-chance JOL accuracy, but elaborative encoding and study practice did not. Notably, although retrieval practice and elaborative encoding led to statistically equivalent final memory performance, they did not lead to the same levels of JOL accuracy. This suggests that the effects of these study techniques on JOL accuracy are dissociable from their effects on long-term retention (cf. Shaughnessy and Zechmeister 1992). That is, more effective study techniques do not necessarily lead to higher JOL accuracy than less effective study techniques.

The average JOLs of participants in the retrieval practice and elaborative encoding groups did not reflect the large advantage of these techniques over study practice on final test performance. These results suggest that children did not monitor the relative efficacy of study techniques on memory retention. This finding coheres with research indicating that the ability to monitor the relative efficacy of study techniques increases through adolescence and into adulthood (cf. Pressley et al. 1984; Beuhring and Kee 1987a, b).

Experiment 2

The results of Experiment 1 suggest that study techniques may influence JOL accuracy in children; however, although we report a medium effect size (as per Cohen 1988), the statistical significance was modest. Experiment 2 examined whether college-aged adults would demonstrate a similar pattern as found with children in Experiment 1. Extending this investigation to college-aged adults is important because extant research has not yet demonstrated an influence of study techniques on delayed-JOL accuracy; however, ceiling-levels of JOL accuracy in these earlier studies may account for the null results (cf. Koriat et al. 2006; Jang et al. 2012). As in Experiment 1, we used a four-alternative forced-choice test to measure memory performance, rather than the more common cued-recall test, to reduce the likelihood of ceiling effects on JOL accuracy. We expected that retrieval practice would improve JOL accuracy in young adults as compared to study practice. We did not make a hypothesis regarding the influence of elaborative encoding on JOL accuracy relative to the other study techniques because to the best of our knowledge, no study has demonstrated that elaborative study techniques lead to different levels of JOL accuracy compared to other study techniques.

Importantly, Experiment 2 was not a strict replication of Experiment 1 with an adult population. Pilot testing with adults using the same methodology of Experiment 1 resulted in ceiling performance in all groups. To address this issue, we increased the retention interval between initial study and JOLs to 48 h and the number of word pairs from 16 to 26. Conducting a follow-up experiment with young adults also allowed us to control for differences in exposure time to the cue words across study technique groups in Experiment 1, which was a limitation of the design. That is, in Experiment 1, learners in the study practice and elaborative encoding groups were exposed to cue words for a fixed interval, whereas participants in the retrieval practice group viewed the cue until a response was entered. In Experiment 2, we ensured that participants in all study technique groups were exposed to each cue-word for a fixed interval.

Participants

Ninety-Six Tufts undergraduate students (24 males, 72 females) aged 18 to 25 ($M_{age} = 19.44$, $SD = 1.46$) participated for course credit or $10 per hour. We randomly assigned 32 participants to the study practice group, 33 participants to the elaborative encoding group, and 31 participants to the retrieval practice group.

Materials

Study material consisted of twenty-nine weakly-related, cue-target word pairs. The mean forward associative strength of the cue to target was 3% ($SD = 2.4\%$) per the University of South Florida Free Association norms (Nelson et al. 1998). These norms were also used to acquire three weakly-related words used as foils for each word pair on the final associative recognition test. The mean of forward associative strength from cue to foil was 5% ($SD = 12\%$).

Design

We used a between-subjects design consisting of three study technique groups: study practice, elaborative encoding, and retrieval practice.

Procedure

As outlined above, the procedure of Experiment 2 matched Experiment 1 except for three changes. First, we increased the Phase 3 retention interval from 15 min to 48 h. Second, we increased the number of word pairs from 16 to 26. Finally, we equated exposure to each cue-word across groups. Participants in all groups were exposed to each cue-word for 2500 ms total per item. Participants began by completing a short practice session in which they practiced each experimental phase with three word-pairs. Presentation of items was randomized across all participants in all phases.

Phase 1. Participants studied all 26 word-pairs for 1000 ms each.
Phase 2. Phase 2 immediately followed Phase 1 and differed across the three groups.

Study practice Participants in the study practice group studied all 26 word-pairs for 1500 ms each.

Elaborative encoding Participants viewed all 26 word-pairs for 1500 each. After a given word pair disappeared, participants were asked to enter in their mediator word for that word pair. On average, it took participants 4753 ms ($SD = 2610$ ms) to enter a mediator word. Participants in the elaborative encoding group left .5% of responses blank.

Retrieval practice Participants viewed each cue-word for 1000 ms. After each cue-word disappeared, participants were asked to enter in the corresponding target word. There was no time limit to respond. After entering a response, participants saw the intact word pair for 500 ms as feedback. On average, it took participants 2993 ms ($SD = 1046$ ms) to produce a response. Mean accuracy during retrieval practice was 33% (SD = 17%) and participants left 3.8% of responses blank.

Phase 3 and 4: JOL and final test Participants then made self-paced JOLs after the 48-h retention interval, in which they were presented each cue one at a time and made their prediction on a 0–10 scale. Average time to enter a JOL was 2736 ms (SD = 877 ms) in the study practice group, 2946 ms (SD = 936 ms) in the elaborative encoding group, and 3248 ms (SD = 1025 ms) in the retrieval practice group. Finally, participants took the self-paced, four-alternative forced-choice test. Average time to enter a response on each test item was 4746 ms (SD = 1277 ms) in the study practice group, 4702 ms (SD = 1297 ms) in the elaborative encoding group, and 3862 ms (SD = 1360 ms) in the retrieval practice group.

Results

Final test performance and mean JOL magnitude

A one-way between-subjects ANOVA on average final test performance revealed a main effect of group, F (2, 87) = 61.83, $p < .001$, $\eta_p^2 = .55$. As shown in Table 2, retrieval practice (M = .79) resulted in higher test performance than study practice (M = .47), t (62) = 9.80, $p < .001$, d = 2.45, as did elaborative encoding (M = .70), t (64) = 7.19, $p < .001$, d = 1.93. Retrieval practice also led to higher test performance than elaborative encoding, t (63) = 2.74, p = .022, d = 0.67.

A one-way between-subjects ANOVA on mean JOL magnitude revealed a main effect of group, F (2, 87) = 7.33, p = .001, $\eta_p^2 = .168$. As shown in Table 2, retrieval practice (M = 5.08) resulted in higher JOL magnitude than study practice (M = 3.71), t (62) = 3.86, p = .001, d = 0.91, as did elaborative encoding (M = 4.98), t (64) = 3.65, p = .001, d = 0.94. No other differences were significant.

JOL accuracy

To measure JOL accuracy, we calculated intra-individual γ correlations between each participant's JOLs and test accuracy on an item-by-item basis. A one-way between-subjects ANOVA on average JOL accuracy revealed a main effect of group, F (2, 89) = 7.04, p = .001, $\eta_p^2 = .14$. We could not compute γ correlations for three participants from the retrieval practice group and one for the elaborative encoding group due to lack of variation

Table 2 Experiment 2. Results of adults. Mean JOLs (for all items and as a function of memory success on the final test), Final Test Performance, and JOL Accuracy (mean γ correlation coefficient)

Study Technique Group	JOL (All)	JOL (Incorrect)	JOL (Correct)	Final Test	JOL Accuracy
Study Practice	3.71 (*1.48*)	3.43 (*1.49*)	3.99 (*1.63*)	.47 (*.12*)	.16 (*.28*)
Elaborative Encoding	4.98 (*1.20*)	4.12 (*1.22*)	5.34 (*1.25*)	.70 (*.12*)	.41 (*.31*)
Retrieval Practice	5.08 (*1.51*)	3.84 (*1.69*)	5.38 (*1.44*)	.79 (*.15*)	.37 (*.26*)

Standard deviations given in parentheses in JOLs or test performance, and thus these four participants were excluded from the analysis. As shown in Fig. 2, retrieval practice ($M = .37$) resulted in higher JOL accuracy than study practice ($M = .16$), t (59) = 2.87, $p = .015$, $d = 0.78$, as did elaborative encoding ($M = .41$), t (59) = 3.50, $p = .002$, $d = 0.84$. The difference in JOL accuracy between retrieval practice and elaborative encoding was not significant ($p > .05$).

We conducted three one-sample t-tests to assess whether mean γ values for each group were statistically greater than 0, which would indicate above-chance JOL accuracy. The JOL accuracy of learners in the retrieval practice group ($M = .37$) was higher than 0, t (27) = 7.65, $p < .001$, $d = 1.45$, as it was for learners in the elaborative encoding group, t (31) = 7.49, $p < .001$, $d = 1.32$, and study practice group ($M = .16$), $t(31) = 7.49$, $p < .001$, $d = 0.56$.

Experiment 2 discussion

As with children, we found that study techniques differentially influenced the accuracy of delayed JOLs in adults. Both retrieval practice and elaborative encoding led to higher JOL accuracy than study practice. We again found evidence that study techniques can influence JOL accuracy independently of their influence on memory retention. Specifically, retrieval practice led to better final test performance than elaborative encoding, but equivalent JOL accuracy.

Further, unlike children in Experiment 1, adults exhibited metacognitive sensitivity, as measured by average JOLs, to the relative efficacy of study techniques. The average JOLs of learners in the retrieval practice and elaborative encoding groups reflected the large memorial advantage over study practice, which accords with previous findings (see Tullis et al. 2013 and Begg et al. 1989, respectively). However, this metacognitive sensitivity was somewhat coarse in that it did not detect the smaller advantage of retrieval practice over elaborative encoding.

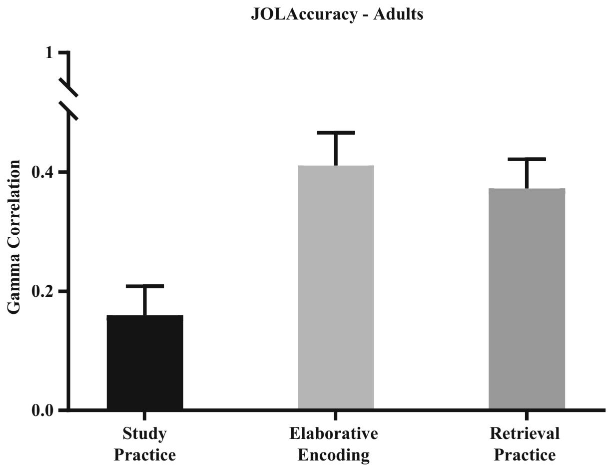

Fig. 2 Delayed judgment-of-learning accuracy of children, as measured by gamma correlation coefficients between JOLs and final test performance on an item-by-item basis. Error bars represent standard error of the mean

General discussion

The present study investigated the factors that influence JOL accuracy in adolescent children and college-aged adults. Consistent with our predictions, we found that study techniques differentially influenced JOL accuracy in both age groups. In children and adults, retrieval practice led to better JOL accuracy than study practice. However, the influence of elaborative encoding on JOL accuracy differed between the age groups. In adults, but not children, elaborative encoding led to higher JOL accuracy than study practice.

Our results also suggest that adults may be better at monitoring the relative efficacy of study techniques than adolescent children. The average JOLs of adults reflected the large memorial advantage of retrieval practice and elaborative encoding over study practice. This accords with previous research demonstrating that adults express higher average delayed JOLs after encoding material through more effective compared to less effective study techniques (Begg et al. 1989; Dunlosky and Nelson 1994, Tullis et al. 2013).

Non-criterial recollection

Our findings provide clues about the types of cues that learners rely on when making delayed JOLs. A great deal of research demonstrates that learners base their JOLs on whether they can retrieve the target (e.g., Nelson et al. 2004; Pyc et al. 2014). That is, learners tend to express high JOLs when they successfully retrieve the target, and low JOLs when they fail to retrieve target. This is also thought to underlie much of the benefit of delay on JOL accuracy because target retrievability from long-term memory is highly diagnostic of later memory performance (Nelson and Dunlosky 1991). If learners relied solely on target retrievability to inform their JOLs, then type of initial study would not influence JOL accuracy. Given that we observed differences in JOL accuracy as a function of how learners initially encoded material, this suggests that (a) learners relied on cues other than target retrievability, and (b) the accessibility of these cues depended on how learners initially studied material.

Our results accord with a growing body of literature demonstrating that *non-criterial* cues influence monitoring judgments. Non-criterial cues refer to any information retrieved about a target that is not the target itself (Parks 2007), such as contextual details from the original encoding event. Research suggests that retrieving such contextual details can improve the accuracy of JOLs (Daniels et al. 2009; McCabe and Soderstrom 2011) and feeling-of-knowing judgments (Brewer et al. 2010; Hertzog et al. 2014; Isingrini et al. 2016; Thomas et al. 2011, 2012), which are predictions about one's ability to recognize currently unrecallable items. Further, how learners encode material influences the accessibility of contextual details (Cook et al. 2006; Hertzog et al. 2014), and thus it is possible that the study techniques we used influenced the accessibility of non-criterial cues. In support of this notion, Hertzog et al. (2014) found not only that retrieving contextual details enhanced feeling-of-knowing accuracy, but also that retrieving contextual details increased with the number of times that learners studied material.

Non-criterial recollection and study techniques

Retrieval practice may increase the quantity and/or accessibility of non-criterial cues relative to study practice. Specifically, research suggests that retrieval practice increases the amount of semantic (Carpenter 2009, 2011; Pyc and Rawson 2010) or contextual information (Lehman et al. 2014; Whiffen and Karpicke 2017) associated with targets. Retrieving these non-criterial cues is thought to facilitate retrieval of the target (Carpenter 2009, 2011; Lehman et al. 2014; Pyc and Rawson 2010; Whiffen and Karpicke 2017). Thus, when prompted to make a JOL, learners who encoded material through retrieval practice may have access to more cues that are predictive of future memory performance than those who encoded via study practice. This would explain the superior JOL accuracy we observed in the retrieval practice groups in both children and adults.

Elaborative encoding involves overtly generating non-criterial cues in the form of mediator words. Research has shown that when prompted to make JOLs, learners are more confident when they retrieve such mediators, even when they cannot retrieve the target (Hertzog et al. 2014). Recollecting these mediators has also been shown to be predictive of later memory performance (Dunlosky et al. 2005; Hertzog et al. 2014). Thus, relative to study practice, elaborative encoding could yield more non-criterial cues that are diagnostic of later memory performance. Our results support this possibility. Elaborative encoding led to the highest JOL accuracy in adults, suggesting that this technique increased the accessibility of non-criterial cues compared to study practice, and that adults effectively capitalized on these cues to predict memory performance. Unlike with adults, our results suggest that children may not have effectively used the non-criterial cues generated during elaborative encoding to predict memory. Research suggests that adolescent learners might not use cues from elaborative tasks effectively because they struggle to assess the quality and completeness of such cues when retrieved from long-term memory (see van Loon et al. 2017).

An alternative to the non-criterial recollection account of our data is that differences in JOL accuracy in our study owed to how study techniques influenced the overall accessibility of targets. If retrieving a target is highly predictive of future memory success, then one might expect that the study techniques that lead to the highest overall levels of target retrievability during the JOL phase would lead to the highest levels of JOL accuracy. However, our results provide evidence against this possibility because we observed dissociations between memory performance and JOL accuracy. In children, retrieval practice and elaborative encoding led to statistically equivalent final test performance. However, equivalent learning, and likely equivalent target retrievability during the delayed-JOL phase, did not result in equivalent levels of JOL accuracy. In addition, elaborative encoding led to higher final test performance than study practice, but did not lead to higher JOL accuracy. Similarly, in adults, retrieval practice led to higher memory performance than elaborative encoding, but statistically equivalent JOL accuracy.

Limitations

One potential confound of our results is differences in study time across groups. However, differences in study time did not appear to influence JOL accuracy significantly in either age-group. Children in the elaborative encoding group spent more time studying than those in the retrieval practice and study practice groups, but exhibited the lowest JOL accuracy. Likewise, although adults in the elaborative encoding group studied longer than those in the retrieval practice group, this did not lead to superior JOL accuracy. Thus, it appears that the type of encoding process, rather than duration of study, accounted for differences in JOL accuracy.

The use of a recognition test to assess final memory performance can complicate the interpretation of JOL accuracy. Judgment-of-learning accuracy is generally higher when memory performance is assessed with cued-recall tests rather than recognition tests (Weaver and Kelemen 2003; Thiede and Dunlosky 1994). This effect can be accounted for in two ways. First, research suggests that learners struggle to predict the extent to which automatic or familiarity-based memory processes, which play a large role on recognition tests, will influence future memory performance (Undorf et al. 2016). Second, correct guessing on recognition tests can suppress JOL accuracy by introducing noise into the measurement of memory performance. Reducing the probability of correct guessing, such as by increasing the number of foils on the recognition test, increases JOL (Thiede and Dunlosky 1994) and feeling-of-knowing (Schwartz and Metcalfe 1994) accuracy. Because we did not include a way to account for correct guessing, we cannot determine the extent to which low levels of JOL accuracy in some groups reflected the difficulty in predicting automatic memory processes versus noise in the measurement of memory performance.

Finally, although our data are consistent with the non-criterial recollection account, our design did not permit a direct test of the account because we did not measure non-criterial recollection. To test the non-criterial recollection account, future research would need to measure non-criterial recollection as a function of the type of encoding task, the degree to which non-criterial recollection is related to future memory performance, and the extent to which learners rely on non-criterial recollection to inform their JOLs.

Conclusion

Study techniques differentially influenced the delayed-JOL accuracy of sixth-grade children and college-aged adults. Our results provide evidence about the types of information learners use to make delayed JOLs. If learners relied only on target-retrievability to make JOLs, we would not have observed differences in JOL accuracy across groups. Our study therefore provides evidence for the hypothesis that learners rely on non-criterial cues to make JOLs, and that the accessibility of these cues depends on how learners encode material. Future research will be needed to test this hypothesis directly.

Acknowledgements The authors would like to thank John Kitzmiller, a mathematics educator at the Frances C. Richmond Middle School at Hanover, NH.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*(5), 610–632.

Beuhring, T., & Kee, D. W. (1987a). Elaborative Propensities During Adolescence: The Relationships Among Memory Knowledge, Strategy Behavior, and Memory Performance. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and Related Mnemonic Processes*. New York: Springer.

Beuhring, T., & Kee, D. W. (1987b). Developmental relationships among metamemory, elaborative strategy use, and associative memory. *Journal of Experimental Child Psychology, 44*(3), 377–400.

Brewer, G. A., Marsh, R. L., Clark-Foos, A., & Meeks, J. T. (2010). Noncriterial recollection influences metacognitive monitoring and control processes. *The Quarterly Journal of Experimental Psychology, 63*(10), 1936–1942.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563–1569.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1547–1552.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2006). Source memory in the absence of successful cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 828–835.

Daniels, K. A., Toth, J. P., & Hertzog, C. (2009). Aging and recollection in the accuracy of judgments of learning. *Psychology and Aging, 24*(2), 494–500.

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*(4), 545–565.

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology, 41*(2), 389–400.

Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction, 32*, 1–9.

Hertzog, C., Fulton, E. K., Sinclair, S. M., & Dunlosky, J. (2014). Recalled aspects of original encoding strategies influence episodic feelings of knowing. *Memory & Cognition, 42*(1), 126–140.

Hoffmann-Biencourt, A., Lockl, K., Schneider, W., Ackerman, R., & Koriat, A. (2010). Self-pace study time as a cue for recall predictions across school age. *British Journal of Developmental Psychology, 28*(4), 767–784.

Isingrini, M., Sacher, M., Perrotin, A., Taconnat, L., Souchay, C., Stoehr, H., & Bouazzaoui, B. (2016). Episodic feeling-of-knowing relies on noncriterial recollection and familiarity: Evidence using an online remember-know procedure. *Consciousness and Cognition: An International Journal, 41*, 31–40.

Jang, Y., Wallsten, T. S., & Huber, D. E. (2012). A stochastic detection and retrieval model for the study of metacognition. *Psychological Review, 119*(1), 186–200.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*(4), 469–486.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology, 93*(2), 329–343.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370.

Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy. In P. Chambres, M. Izaute, & P. Marescaux (Eds.), *Metacognition: Process, function and use* (pp. 1–17). Dordrecht: Kluwer Academic Publishers.

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., Ma'ayan, H., Sheffer, L., & Bjork, R. A. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 595–608.

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The easily learned, easily remembered heuristic in children. *Cognitive Development, 24*(2), 169–182.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(6), 1787–1794.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 464–470.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 756–766.

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition, 18*(2), 196–204.

McCabe, D. P., & Soderstrom, N. C. (2011). Recollection-based prospective metamemory judgments are more accurate than those based on confidence: Judgments of remembering and knowing (JORKS). *Journal of Experimental Psychology: General, 140*(4), 605–621.

Meeter, M., & Nelson, T. O. (2003). Multiple study trials and judgments of learning. *Acta Psychologica, 113*(2), 123–132.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*(3), 159–163.
Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning, 8*(1), 19–46.
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*(4), 267–270.
Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 26, pp. 125–173). New York: Academic Press.
Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25, Chapter xiii, 334 Pages)). Cambridge: The MIT Press.
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida Word Association, Rhyme, and Word Fragment Norms.
Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods, 9*(1), 53–69.
Parks, C. M. (2007). The role of noncriterial recollection in estimating recollection and familiarity. *Journal of Memory and Language, 57*(1), 81–100.
Peynircioğlu, Z. F., Brandler, B. J., Hohman, T. J., & Knutson, N. (2014). Metacognitive judgments in music performance. *Psychology of Music, 42*(5), 748–762.
Pressley, M., Levin, J. R., & Ghatala, E. S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning & Verbal Behavior, 23*(2), 270–288.
Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*(6002), 1–335.
Pyc, M. A., Rawson, K. A., & Aschenbrenner, A. J. (2014). Metacognitive monitoring during criterion learning: When and why are judgments accurate? *Memory & Cognition, 42*(6), 886–897.
Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The oxford handbook of metamemory* (pp. 65–80). New York: Oxford University Press.
Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin, 137*(1), 131–148.
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*(2), 117–137.
Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's monitoring: Evidence from a judgment-of-learning. *Cognitive Development, 15*, 115–134.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Pittsburgh: Psychology Software Tools.
Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93–113, Chapter xiii, 334 Pages)). Cambridge: The MIT Press.
Shaughnessy, J. J., & Zechmeister, E. B. (1992). Memory-monitoring accuracy as influenced by the distribution of retrieval practice. *Bulletin of the Psychonomic Society, 30*(2), 125–128.
Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6*(4), 662–667.
Thiede, K. W., & Dunlosky, J. (1994). Delaying students' metacognitive monitoring improves their accuracy in predicting their recognition performance. *Journal of Educational Psychology, 86*(2), 290–302.
Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73.
Thomas, A. K., & McDaniel, M. A. (2007). Metacomprehension for educationally relevant materials: Dramatic effects of encoding-retrieval interactions. *Psychonomic Bulletin & Review, 14*(2), 212–218.
Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2011). Context affects feeling-of-knowing accuracy in younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 96–108.
Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2012). An analysis of the determinants of the feeling of knowing. *Consciousness and Cognition: An International Journal, 21*(4), 1681–1694.
Tsalas, N., Paulus, M., & Sodian, B. (2015). Developmental changes and the effect of self-generated feedback in metacognitive controlled spacing strategies in 7-year-olds, 10-year-olds, and adults. *Journal of Experimental Child Psychology, 132*, 140–154.
Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition, 41*(3), 429–442.
Undorf, M., Böhm, S., & Cüpper, L. (2016). Do judgments of learning predict automatic influences of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(6), 882–896.

van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. J. G. (2013). The effect of delayed-JOLs and sentence generation on children's monitoring accuracy and regulation of idiom study. *Metacognition and Learning, 8*(2), 173–191.

van Loon, M., de Bruin, A., Leppink, J., & Roebers, C. (2017). Why are children overconfident? developmental differences in the implementation of accessibility cues when judging concept learning. *Journal of Experimental Child Psychology, 158*, 77–94.

von der Linden, N., Löffler, E., & Schneider, W. (2015). Effects of a short strategy training on metacognitive monitoring across the life-span. *Frontline Learning Research, 3*(4), 36–55.

Weaver, C. A., & Kelemen, W. L. (2003). Processing similarity does not improve metamemory: Evidence against transfer-appropriate monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1058–1065.

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1036–1046.