Cognitive Development 24 (2009) 169–182

Contents lists available at ScienceDirect

Cognitive Development

The easily learned, easily remembered heuristic in children

Asher Koriat[a,*], Rakefet Ackerman[a], Kathrin Lockl[b], Wolfgang Schneider[c]

[a] University of Haifa, Haifa, Israel
[b] University of Bamberg, Bamberg, Germany
[c] University of Würzburg, Würzburg, Germany

ARTICLE INFO

Keywords:
Metacognitive development
Learning
Judgments of learning
Metacognitive heuristics
Metamemory

ABSTRACT

A previous study with adults [Koriat, A. (2008a). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, *36*, 416–428] established a correlation between learning and remembering: items requiring more trials to acquisition (TTA) were *less* likely to be recalled than those requiring fewer trials. Furthermore, learners' judgments of learning (JOLs) seemed to rely on the *easily learned*, *easily remembered* (ELER) heuristic, that items requiring fewer TTAs are more likely to be recalled. This study extended investigation of these effects to 2nd- and 4th-grade children. When the list included hard and easy paired-associates (Experiment 1, $N = 40$, 7–10 years), recall and JOL decreased with increasing TTAs for both grades, supporting the validity of the ELER heuristic and its utilization in monitoring one's own learning. When presented only with hard pairs (Experiment 2, $N = 60$, 7–10 years), however, 4th graders' but not 2nd graders' JOLs evidenced reliance on this heuristic. The results suggest an early development of metacognitive heuristics that incorporate information about the links between characteristics of the encoding process and subsequent remembering.

There has been extensive research among children and adults on the metacognitive processes that take place during learning. The assumption underlying that research is that the monitoring of one's own learning and the regulation of learning strategies may have pronounced effects on subsequent memory performance (Benjamin & Bjork, 1996; Brown, 1975; Flavell, 1979; Lucangeli, Galderisi, & Cornoldi, 1995; Schneider, 1999; Schneider & Lockl, 2008; Schneider & Pressley, 1997).

* Corresponding author at: Department of Psychology, University of Haifa, Haifa 31905, Israel. Tel.: +972 4 8249432; fax: +972 4 8249431.
E-mail address: akoriat@research.haifa.ac.il (A. Koriat).

doi:10.1016/j.cogdev.2009.01.001

How do learners monitor their knowledge during study? Underlying much of the work that addressed this question is a cue-utilization view which assumes that learners' judgments of learning (JOLs) are inferential in nature and based on cues and heuristics that have a certain degree of validity in predicting memory performance. Koriat (1997) distinguished three types of cues: intrinsic, extrinsic, and mnemonic. Intrinsic cues refer to inherent characteristics of the study items, such as the degree of associative relatedness between the members of a paired-associate. Extrinsic cues pertain to the conditions of learning (e.g., number of presentations) or to the encoding operations applied by the learner (e.g., level of processing). Finally, mnemonic cues are internal, subjective indicators that signal to the person the extent to which an item has been mastered. These cues reside in the feedback that learners gain from attempting to master the material.

The present study, like most of the recent research on the bases of metacognitive judgments, focuses on the last category of these cues – mnemonic cues. Evidence suggests that JOLs are influenced by the ease with which to-be-remembered items are encoded during learning (Benjamin & Bjork, 1996; Dunlosky & Nelson, 1992; Koriat & Ma'ayan, 2005). For example, JOLs were found to increase with the success and speed of forming an interactive image between the cue and the target during paired-associate learning (Hertzog, Dunlosky, Robinson, & Kidder, 2003). Several results suggest that JOLs also increase with the fluency with which items are retrieved during learning, and that retrieval fluency is generally diagnostic of subsequent recall performance (Koriat & Ma'ayan, 2005; Nelson, Narens, & Dunlosky, 2004; Son & Metcalfe, 2005). However, recall predictions have been found to increase with the speed of retrieving an answer even under conditions in which actual recall exhibited the opposite effect (Benjamin, Bjork, & Schwartz, 1998).

The cue-utilization view calls for an analysis similar to that implied by Brunswik's (1956) lens model for visual perception. According to that model, several proximal cues in the retinal image are used automatically and unconsciously to produce a sheer perceptual experience. The accuracy of perception in capturing the properties of distal objects (e.g., actual size) depends on the validity of the various proximal cues available (cue validity) and on the extent to which these cues are utilized in inferring distal properties (cue utilization). In like manner, a systematic study of metacognitive judgments during learning and their accuracy in predicting memory performance requires examination of the following: (a) the validity of the cues that are used by the learner to infer future memory performance (cue validity), (b) the extent to which these cues are used by the learner in making JOLs (cue utilization), and (c) the resultant overall accuracy of metacognitive judgments in predicting memory performance (achievement).

The idea that JOLs are based on the fluency with which items are encoded during study has several implications that seem counterintuitive at first sight. These implications are illustrated by two recent studies, the first concerning self-paced study time (Koriat, Ma'ayan, & Nussinson, 2006) and the second the number of trials to acquisition (TTA) (Koriat, 2008a). In the first of these, Koriat et al. (2006) asked college students to study a list of paired-associates under self-paced conditions. JOLs were found to be inversely related to study time. That is, the more time learners invested in the study of an item, the *lower* were their predictions that the item would be recalled in the future. This inverse relationship contrasts with what is typically found when study time is experimentally manipulated. In that case, JOLs *increase* with presentation duration (Koriat, 1997). These correlations mirror the within-person JOL–recall correlations: in self-paced learning, recall *decreases* with the amount of study time invested during learning, but it *increases* with presentation time when presentation time is manipulated experimentally between items. For example, Bjork (1999) observed that manipulations that create difficulties for the learners tend to benefit memory because they induce learners to engage in effective processes that they would not have used otherwise.

The contrast between the situation in which JOLs decrease with study time and that in which they increase with study time could be seen to reflect the difference between self-regulation and other-regulation. Koriat et al. (2006), however, showed that even when the allocation of study time is regulated by the learner, JOLs and recall *increase* with study time when the self-regulation of study time is goal-driven. When different incentives were awarded to the recall of different items within a list, learners devoted more study time to the high-incentive than to the low-incentive items and in parallel assigned higher JOLs to the high-incentive than to the low-incentive items. Thus, according to Koriat et al. (2006), the critical contrast in self-paced learning is between goal-driven regulation and data-driven regulation. When learners allocate study time according to some goal, for example spending more time studying for an exam on a topic that is particularly important for them than for other topics, JOLs (as well as recall) are expected to increase with study time. In contrast, in the typical self-paced learning situation, the allocation of study time between same-incentive items is largely data-driven. In other words, the amount of time invested in an item is determined ad hoc by the item itself, or rather, by the item–learner interaction. Once a learner has managed to commit an item to memory, for example by discovering or forming an association between the cue and the target, the learner uses study time (or memorizing effort) as a cue for JOLs under the heuristic that easily learned items are more likely to be remembered than items that require more effort to learn (Koriat, 2008a; Koriat & Ma'ayan, 2005). The inverse relationship between JOLs and study time, although counterintuitive, is consistent with the idea that JOLs are based on the fluency with which items are encoded or retrieved during study (Benjamin & Bjork, 1996).

Koriat et al. (2006) argued that the inverse relationship between JOLs and study time is the signature of data-driven regulation, in which monitoring follows control rather than preceding it. This view departs from previous conceptualizations in which study time allocation in self-paced learning is assumed to be goal-driven: learners invest relatively more study time in the judged-difficult items in order to compensate for their difficulty (e.g., Nelson & Leonesio, 1988). By this account, monitoring (JOL or judged ease of learning) drives control (study time allocation). According to Koriat et al. (2006), in contrast, it is the amount of effort invested in an item that informs the learner about the difficulty of the item. Thus, metacognitive judgments are sometimes based on the feedback from control processes. A similar process seems to underlie confidence judgments: confidence in the correctness of an answer or a solution *decreases* with the time it takes to reach that answer or solution. Presumably learners spend as much time as is needed to retrieve an answer or reach a solution (control) and then use retrieval time (or effort) as one of the cues for their confidence (Kelley & Lindsay, 1993; Koriat, 2008b).

A second study (Koriat, 2008a) that focused on number of trials to acquisition also brings to the fore the unique characteristics of data-driven regulation. College students studied a list of paired-associates for several study–test cycles under fixed-rate presentation. A dropout procedure was used so that each item recalled on a particular cycle was deleted from subsequent study–test cycles. A final cued recall test of all items took place 7 min after the last cycle (in which all the remaining items were recalled correctly). The results yielded an *inverse* relationship between trials to acquisition and final recall. These results were described in terms of the rule *easily learned*, *easily remembered* (ELER). Additional results indicated that metacognitive judgments incorporate knowledge of the ELER rule. When learners made JOLs regarding their performance in the final cued–recall test, they expected better recall for items that required *fewer* trials to acquisition than for those requiring more trials. The ELER principle is the opposite of what is implied by the adage "easy comes, easy goes", but is consistent with the general idea that when learning is data-driven, determined by the item itself, ease of learning (EOL) is predictive of remembering, such that easily encoded information is more likely to be recalled later than is information that requires more effort to encode. This general principle seems to be incorporated in metacognitive judgments, but it is not clear whether the cues underlying the relationship between JOL and TTA are the same as those underlying the relationship between JOL and self-paced study time.

The present study is part of a project that attempts to explore the data-driven regulation of study within a developmental perspective. Whereas a previous study (Koriat, Ackerman, Lockl & Schneider, 2009) examined the relationship between JOLs and study time in self-paced learning, the present study focuses on the relationships between TTA, on the one hand, and JOL and recall, on the other. In the Koriat et al. (2009) study, 3rd to 6th graders studied a list of paired-associates under self-paced instructions. Their JOLs were found to be inversely related to the amount of time invested in each item, similar to what had been observed for young adults (Koriat et al., 2006). The results for 1st and 2nd graders, in contrast, did not yield this relationship, suggesting that young children do not use the feedback from study experience as a cue for recall predictions. Note that as far as cue validity is concerned, study time was predictive of actual recall even for the younger children, suggesting that these children fail to take advantage of a cue that is predictive of subsequent memory performance.

In the present study we extended investigation of the ELER heuristic by examining how JOLs and recall vary with TTAs for 2nd and 4th graders. Three questions were addressed. The first concerns the actual correlation between learning and remembering (cue validity): at what age does speed of learning become a valid predictor of future recall? For example, would 2nd graders also exhibit a *negative* correlation between TTA and subsequent recall? This question concerns what Koriat (2008a) referred to as the "internal ecology" of cognitive processes. The second question concerns cue utilization. Would children's JOLs also disclose the heuristic that the more trials they spend studying an item until acquisition the *less* likely they are to recall that item? The third question concerns self-awareness. Could it be the case that children's metacognitive judgments disclose reliance on this somewhat counterintuitive heuristic even when they cannot state explicitly the ELER rule?

The present study departs from much of the work on children's metacognitive skills, which has placed greater emphasis on children's declarative knowledge (Schneider & Bjorklund, 1998; Schneider & Lockl, 2008). Here, in contrast, we focus on the heuristics that underlie metacognitive judgments. These heuristics may be based on mnemonic cues that reside in the feedback from task performance and may operate below full consciousness (Koriat, 2007). A few recent studies have explored other heuristics that underlie metacognitive judgments in children. For example, Lockl and Schneider (2002), who examined children's feelings of knowing (FOK), found evidence suggesting that children, like adults, rely on the accessibility heuristic (Koriat, 1993): they base their FOK judgments on the amount of information that can be accessed regardless of its correctness. Similarly, results by Roebers, von der Linden, Schneider, and Howie (2007) suggest that children's JOLs also rely in part on the amount of information that comes to mind and its ease of access. Ghetti and her associates examined the development of the memorability heuristic for rejecting false events (Ghetti, 2003; see Ghetti, 2008, for a review). Whereas adults tend to be more confident that an event has not occurred if that event is highly memorable than if it is not so memorable, children under 9 years fail to consider the memorability of the event in judging its nonoccurrence. Further results suggest that this failure does not derive from their inability to appreciate the memorability of an event but from specific metacognitive limitations to consider expected memorability when making decisions about past occurrences.

1. Study 1

In this study, 2nd and 4th graders studied a list of items for several study–test cycles, and items recalled in one cycle were deleted from further study and test. All pairs were then presented, and participants indicated their JOLs about each item. A final cued–recall phase took place. We examined whether items studied for more trials are indeed less likely to be recalled than those studied for fewer trials (cue validity), and whether the ELER heuristic underlies learners' JOLs in each of the age groups (cue utilization). A self-report questionnaire was also included to examine whether children can explicitly state the ELER rule and whether they can discriminate between the items that require a few and those that require many TTAs.

1.1. Method

1.1.1. Participants

Forty children (17 females) from elementary schools in Israel participated in the study. The children were mostly of middle-class and upper-middle-class socioeconomic background. The younger age group consisted of 20 2nd graders (mean age 7.3, range 6.8–7.8) and the older age group consisted of 20 4th graders (mean age 9.4, range 8.8–10.1).

1.1.2. Materials

The items were 24 pairs of Hebrew words used in previous research (Koriat et al., 2009). They had been selected on the basis of a norming study. Twelve represented easy pairs; in each of them the words were semantically or associatively related (e.g., *king-crown*). In the remaining (hard) pairs, the two members of each pair were unrelated (e.g., *cake-rug*).

1.1.3. Procedure

Children participated individually in a quiet room in the school, using a laptop computer. They were told that they should study pairs appearing on the computer screen so that, during the test phase, they would be able to recall the response word when cued with the stimulus word. Each pair was presented for 5 s. When the pair disappeared, the child had to press a button when ready to study the next pair. Two practice items were used to familiarize the children with this procedure.

In the test phase that followed, each of the stimulus words were presented in turn (in a different order than in the study phase), and the recalled responses were spoken orally by the child. Children were encouraged to try to recall the response word, but when unable to produce a response, they could continue to the next cue word. The first two cue words were from the practice items.

Only those pairs that children had failed to recall correctly in the preceding test phase were presented for study in the subsequent trial and then presented at test (Karpicke & Roediger, 2008). Children were informed that some of the word pairs would appear again for study (without explaining the principle underlying the selection of items for restudy). The study phase ended when children achieved perfect recall.

Next, the entire list was presented for JOL solicitation. The pairs were presented one after the other for 3 s each, and the child indicated a JOL regarding the likelihood of recalling the target word in response to the cue word in a subsequent cued–recall test. The elicitation of JOLs capitalized on the hot–cold game familiar to children, using a thermometer procedure (Koriat et al., 2009). The rules of the hot–cold game were explained. A colored scale, in the shape of a horizontal thermometer, was presented, with a pointer positioned at the middle of the scale. Children were asked to make their rating by sliding the pointer using the mouse. The position of the pointer on the scale was transformed into a JOL percentage score, with one end of the scale (deep blue, "very cold") defined as 0% and the other (deep red, "very hot") as 100%.

A final test phase took place, using the same cued–recall procedure as before, but including all the word pairs. This final recall was used as a criterion for successful memory performance. The first two cue words were from the practice items. Immediately after these items, as well as at the end of the testing phase, children were asked to make aggregate-JOLs expressing their estimate of the number of items that will be or were correctly recalled. Results for these aggregate-JOL estimates are not reported in this article. Finally, children were asked several questions about the basis of their JOLs. The entire session lasted on average approximately 30 min.

Children were contacted 1–8 weeks later and asked for EOL ratings of the pairs used in the study. A printed form, listing all 24 word pairs was presented, and children were asked to estimate for each pair how many study cycles would be needed to commit that item to memory. A scale from 1 to 10 appeared next to each pair, and children had to circle the number of study trials needed to successfully recall the target word when the cue word is presented.

1.2. Results

Fig. 1 presents mean JOL and mean final recall as a function of number of TTAs for 4th graders (top panel) and 2nd graders (bottom panel). Presented also are the number of items on which each mean is based. Across all participants and items, there were 13 cases in which a child required more than seven trials to acquisition. These observations were eliminated from the figure and from all the analyses that follow.

Preliminary analyses indicated that 2nd graders needed marginally more TTAs per item (2.4) than 4th graders (2.1), $t(38) = 1.95$, $p < .06$. Otherwise there was little difference between the two grades. We nevertheless examine performance for each grade separately in order to assess the reality of the ELER principle for each age group and also to allow comparison with the results of Study 2, which yielded age differences.

1.2.1. Cue validity: final recall as a function of number of trials to acquisition

The results (Fig. 1) indicated that final recall decreased monotonically with TTA for both 2nd and 4th graders, similar to what was found for young adults (Koriat, 2008a). Traditional statistical procedures cannot be applied to the results presented in Fig. 1 because of the "fragmentary data problem" noted in connection with research on the tip-of-the-tongue phenomenon (Brown, 2008): each data point in the figure is based on a different combination of participants and items. However, to convey a rough estimate of the drop in recall that occurred with increasing TTAs, it can be noted that for 4th graders

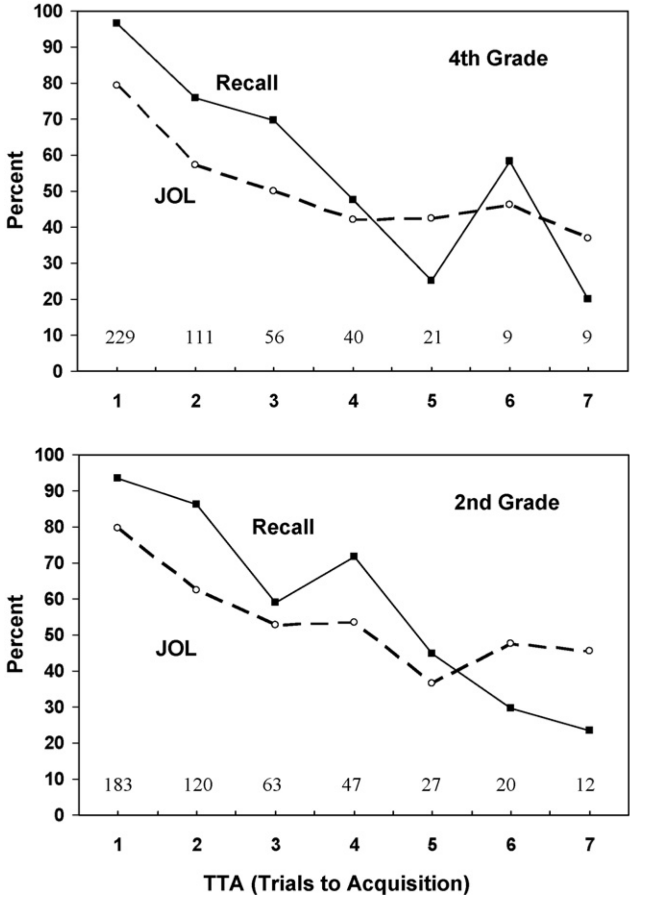

Fig. 1. Percent final recall and JOL in Study 1 as a function of TTA (number of trials to acquisition) for 4th graders (top panel) and for 2nd graders (bottom panel). The bottom line indicates the number of observations on which each mean is based.

recall for items requiring 2 TTAs or less (67.9% of all observations) averaged 91.3%, compared to 56.4% for items requiring 3 TTAs or more. The respective means for the 2nd graders were 92.4% and 55.5%. These results suggest a marked drop in recall with increasing TTA, with a similar effect for the two age groups.

To evaluate the significance of this drop and also to explore possible developmental trends that may be masked when data are aggregated across all participants (Cohen, Sanborn, & Shiffrin, 2008), we performed within-person linear regression analyses using TTA as the predictor and percent recall as the dependent variable. The slope of the regression was negative for all participants in both grades. The slope for 4th graders averaged −13.26, and was significantly different from 0, $t(19) = 7.89$, $p < .0001$. The slope for 2nd graders averaged −12.83 and was also significantly different from 0, $t(19) = 8.12$, $p < .0001$. The difference in mean slope between the two grades was not significant, $t(38) = 0.18$, $p = .86$. Altogether, these results suggest that each additional TTA is associated with a 13% drop in final recall.

These analyses were supplemented by within-person gamma correlations between TTA and final recall. These correlations averaged −.74 for 4th graders, significantly different from 0, $t(19) = 15.45$, $p < .0001$, and −.64, $t(19) = 11.15$, $p < .0001$, for 2nd graders. The correlation was negative for all 4th graders as well as for all 2nd graders, and the difference in mean correlation between the two grades was not significant, $t(38) = 1.34$, $p = .19$.

In sum, the results yielded clear evidence for the ELER rule. Although all items attained perfect recall by the end of learning, the learning history of each item predicted its forgetting (Karpicke & Roediger, 2008). For both grades, final recall decreased as the number of trials to acquisition increased, and there was little difference between the two age groups in this respect. It is thus quite remarkable that even for 2nd graders ease of learning was a powerful predictor of recall success.

1.2.2. Cue utilization: JOLs as a function of number of trials to acquisition

We turn next to metacognitive judgments. The results (Fig. 1) indicate that JOLs also decreased with increasing TTAs. For 4th graders, JOLs were higher for items requiring 1 or 2 presentations (72.6%) than for those requiring 3 or more presentations (43.9%). The respective values for 2nd graders were 74.8% and 48.5%.

Similar within-person regression analyses as those used for recall confirmed that JOLs also decreased significantly with TTA. Thus, the slopes of the regression lines relating JOLs to TTA were negative for all participants. The slopes for the 4th graders (mean −13.41) as well as those for the 2nd graders (mean −11.36) were significantly different from 0, $t(19) = 8.70$, $p < .0001$, and $t(19) = 5.25$, $p < .0001$, respectively. The difference between the two means was not significant, $t(38) = 0.77$, $p = .45$. This pattern of results was also reflected in the JOL–TTA gamma correlations. These correlations averaged −.60 for 4th graders, $t(19) = 16.75$, $p < .0001$, and −.56 for 2nd graders, $t(19) = 18.79$, $p < .0001$. The gamma correlations were negative for all 4th graders as well as for all 2nd graders, and the difference between grades was not significant, $t(38) = 0.91$, $p = .37$. Thus, like recall, JOLs also decreased as the number of trials to acquisition increased, and there was little difference between the two groups in this respect.

We also examined whether the two age groups differed in the degree to which they relied on intrinsic item difficulty, that is, on the contrast between related and unrelated pairs. The within-person gamma correlation between JOLs and item difficulty (scored dichotomously as 1 = related and 2 = unrelated) averaged −.76 for 2nd graders and −.77 for 4th graders with no difference between them, $t(38) = 0.25$, *ns*. Regression analyses also failed to yield a significant difference between the two age groups. Nevertheless, when both TTA and item difficulty were used in multiple regression analysis to predict JOLs, a two-way ANOVA on the standardized beta weights indicated significant effects for predictor and for grade, but not for the interaction. Beta for item difficulty and TTA averaged −.48 and −.26, respectively, $F(1, 38) = 7.17$, $p = .01$. Beta for 2nd and 4th graders averaged −.34 and −.40, respectively, $F(1, 38) = 5.05$, $p < .05$. Thus, there is some indication of a slightly stronger cue utilization by 4th graders than by 2nd graders, but not in the relative weight of the two cues.

1.2.3. Achievement: the accuracy of JOLs in predicting recall performance

Correspondence between cue utilization and cue validity – the observation that JOLs bear the same relationship to TTA as does recall – may be expected to contribute to the accuracy of JOLs in predicting recall. Indeed, the within-person gamma correlations between JOLs and final recall (Nelson, 1984) averaged .56 for 4th graders, significantly different from 0, $t(19) = 8.60$, $p < .0001$. The respective correlation for the 2nd graders was .68, $t(19) = 13.08$, $p < .0001$. The difference between grades was not significant, $t(38) = 1.41$, $p = .17$. Thus, children in both age groups were generally accurate in discriminating between items that were recalled and those that were not recalled.

1.2.4. Self-report data

The self-report data provides some clues regarding the bases of metacognitive judgments. When asked whether they figured out which items were presented for restudy, 60% of the children in each age group indicated that the repeated items were those that had not been recalled correctly in the previous test. Children were also asked to indicate the basis for their JOLs. Overall, 8 children in each grade (40%) thought that more study cycles led them to provide lower JOLs, whereas 10 children in each grade (50%) thought that more study cycles actually led them to give *higher* JOLs (10% gave other responses). Thus, even after having completed the study, children did not evidence awareness of the ELER heuristic. We examined whether the JOL–TTA gamma correlation differed for the children who held different metacognitive beliefs. For those who reported that JOLs increased with TTA, the JOL–TTA correlation averaged −.61 and −.63 for the 4th and 2nd graders, respectively. The respective correlations for those who reported that JOLs decreased with TTA averaged −.56, and −.52. These results suggest that the TTA–JOL correlation was not simply based on an explicit metacognitive inference.

The analysis of the EOL ratings, in contrast, indicated that children were somewhat successful in estimating the number of study cycles needed to master different items. The within-participant Pearson correlation between actual and estimated number of TTAs averaged .43 for 2nd graders and .44 for 4th graders, both $p < .0001$.

2. Study 2

In Study 1, the paired-associates represented different degrees of cue–target associations so that some items were relatively easy to learn whereas others were more difficult. Thus, it is possible, for example, that children's JOLs were based entirely on the normative difficulty of the items, and that normatively difficult items require more TTAs and are also easier to recall. Indeed, previous studies indicated that even 1st and 2nd graders are sensitive to the difference between related and unrelated paired-associates (Dufresne & Kobasigawa, 1989) and assign higher JOLs to the related pairs (Koriat et al., 2009). To examine the question of whether the JOL–TTA relationship for young adults is mediated entirely by inter-item differences in intrinsic difficulty, Koriat (2008a) used only unrelated paired-associates. The results still yielded support for the ELER principle, even though judges failed to estimate accurately the number of TTAs associated with different items. In Study 2, we included only unrelated word pairs so that there was little systematic variation between the pairs in the strength of the a priori association. We examined whether differences between the pairs in number of TTAs, however small and idiosyncratic they may be, will be predictive of subsequent recall, and will also affect JOLs.

2.1. Method

2.1.1. Participants

Sixty children (34 females) from the same schools as in Study 1 participated. The younger group consisted of 30 2nd graders (mean age 7.8, range 7.0–8.6) and the older age group consisted of 30 4th graders (mean age 9.6, range 8.9–10.8).

2.1.2. Materials and procedure

Fifteen unrelated pairs were used. These included the 12 difficult pairs from Study 1, and three additional unrelated pairs taken from the same set used for selecting the materials for Study 1. All 15 pairs were used for the 4th graders, but only 12 of these were used for the 2nd graders to avoid too many study trials for this group. Two additional pairs were used for practice. The procedure was otherwise identical to that of Study 1.

2.2. Results

Fig. 2 presents mean JOL and final recall as a function of TTAs for 4th graders (top panel) and 2nd graders (bottom panel). Presented also are the number of items on which each mean is based. Across all participants and items there were 22 cases in which a child required more than 10 trials to acquisition. These observations were eliminated from the figure and from all subsequent analyses.

Although the study lists differed in length for the two age groups, 2nd graders needed on average more TTAs per item (4.3) than did 4th graders (3.5) and exhibited poorer recall overall (47.4%) than 4th graders (66.2%). JOLs averaged 52.9% for 2nd graders and 58.6% for 4th graders.

2.2.1. Cue validity: final recall as a function of number of trials to acquisition

We compared mean recall for items requiring 3 TTAs or less (54% of all observations) with those requiring 4 TTAs or more. For 4th graders, mean final recall was higher for items requiring 3 or less presentations (74.3%) than for items requiring 4 or more presentations (47.2%; based on 27 children, because 3 children needed 3 TTAs or less for all items). For the 2nd graders, the respective means were 50.8% and 40.8% (based on 29 children, because one child needed more than 3 TTAs for all items).

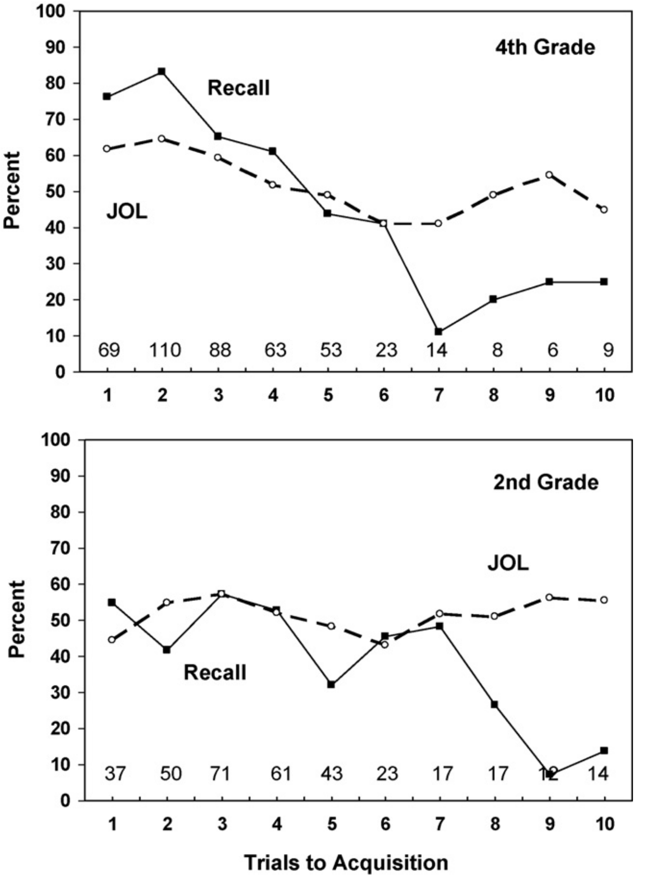

Fig. 2. Percent final recall and JOL in Study 2 as a function of TTA (number of trials to acquisition) for 4th graders (top panel) and for 2nd graders (bottom panel). The bottom line indicates the number of observations on which each mean is based.

Within-person linear regression analyses were carried out as in Study 1, using TTA as a predictor of percent recall. Two 4th graders were eliminated from the analysis because they exhibited 100% final recall. The slope of the regression lines was negative for 22 of the 28 4th graders ($p < .001$, binomial test) and for 22 of 30 2nd graders ($p < .05$, binomial test). The slope for 4th graders averaged −9.12, and was significantly different from 0, $t(27) = 4.13$, $p < .001$. The slope for 2nd graders averaged −5.99 and was also significantly different from 0, $t(29) = 3.24$, $p < .01$. The difference in mean slope between the two grades was not significant, $t(56) = 1.09$, $p < .28$.

The within-person gamma correlation between TTA and final recall averaged −.37 for the 4th graders, $t(27) = 4.37$, $p < .0005$, and was negative for 22 of the 28 children ($p < .005$ by a binomial test). For the 2nd graders it averaged −.31, $t(29) = 4.07$, $p < .0005$. The correlation was negative for 23 of the 30 children ($p < .005$, binomial test). The difference between grades was not significant, $t(56) = 0.50$, $p = .62$. In sum, the results for both grades yielded the same evidence for the ELER relationship as was the case in Study 1.

2.2.2. Cue utilization: JOLs as a function of number of trials to acquisition

Turning next to metacognitive judgments, the results for 4th graders indicated that JOLs also decreased systematically with TTA as was found to be the case in Study 1: JOLs were higher for items requiring 3 or fewer TTAs (60.9%) than for those requiring 4 or more TTAs (50.4%). The respective values for 2nd graders were 56.1% and 51.9%.

The within-person analyses indicated the possibility of an age difference in JOL. The slopes of the regression lines relating JOLs to TTA were negative for 25 of the 30 4th graders ($p < .001$, binomial test) but for only 15 of the 30 2nd graders. The slopes for the 4th graders (mean −4.03) were significantly different from 0, $t(29) = 4.01$, $p < .0005$, whereas those for the 2nd graders (mean −0.32) were not, $t(29) = .26$, $p = .79$. The difference between the two means was significant, $t(58) = 2.32$, $p < .05$.

This age difference was also supported by the within-person TTA–JOL gamma correlations. For 4th graders, these correlations averaged −.22, $t(29) = 4.07$, $p < .0005$, and were negative for 24 of the 30 children ($p < .001$, binomial test). For the 2nd graders, in contrast, the correlations averaged −.04, $t(29) = 0.77$, $p = .45$, and were negative for 15 of the 30 children. The difference between grades was significant, $t(58) = 2.24$, $p < .05$. In sum, JOL decreased significantly with increasing TTA for 4th graders but not for 2nd graders.

The results of Study 2 suggested that the effects of TTA were weaker for JOLs than for recall (Fig. 2). A two-way ANOVA, with Grade and Measure (recall vs. JOL) as factors, on the within-person slopes yielded significant effects for measure, $F(1, 56) = 12.06$, $MSE = 69{,}89$, $p = .001$, and for grade, $F(1, 58) = 4.01$, $MSE = 85.72$, $p < .05$, with no interaction. The effects of TTA were somewhat weaker overall for 2nd graders than for 4th graders, but JOLs decreased more shallowly with TTA than did recall for both grades.

2.2.3. Achievement: the accuracy of JOLs in predicting recall performance

As would be expected, the JOL–recall gamma correlations were lower than those obtained in Study 1, in which there was a greater inter-item variation in intrinsic item difficulty. These correlations averaged .35 for 4th graders and .25 for 2nd graders. Both were significantly different from 0, $t(27) = 4.70$, $p < .0001$, and $t(29) = 3.11$, $p < .005$, and did not differ significantly from each other, $t(56) = 0.99$, $p = .34$. For 4th graders, the correlation was positive for 21 of the 28 children ($p < .01$, binomial test), and for 2nd graders it was positive for 19 of the 30 children ($p < .11$, binomial test). Thus, recall predictions were relatively accurate even though the word pairs were normatively unrelated.

2.2.4. Self-report data

In this study 63% of 2nd graders and 77% of 4th graders identified that items were presented again if they were not recalled correctly in the previous cycle. In reporting on the basis for their JOLs, 43% of 4th graders and 30% of 2nd graders thought that more study cycles led them to provide lower JOLs, whereas 47% of both 4th and 2nd graders thought that more cycles actually led them to give *higher* JOLs (10% and 23% of 4th and 2nd graders gave other responses). Thus, as in Study 1, there was no clear evidence that children were aware of the ELER principle. We examined whether the JOL–TTA gamma correlation differed for the two groups. For the former group, this correlation averaged −.39 and −.22, respectively. The respective correlations for the latter group averaged −.08, and .05, respectively. Thus, there was some consistency between the children's stated beliefs and the actual JOLs exhibited by the child, but the cause-and-effect relationship underlying that consistency is unknown.

Analysis of the EOL ratings indicated that children were not successful in estimating the number of study cycles needed to master different items. The within-participant Pearson correlation between actual and estimated number of TTAs averaged .03 for 4th graders and .04 for 2nd graders. In the study by Koriat (2008a) with college students, judges were asked to estimate the number of TTAs needed to master each item. The correlation between their estimates and the actual TTAs exhibited by participants averaged .03. Thus, it appears that neither external judges nor the learners themselves are successful in estimating the number of repetitions required for acquisition when the pairs are unrelated.

3. Individual differences

Does the inverse relationship between TTA and recall hold true with respect to individual differences? That is, do individuals who require fewer TTAs overall exhibit better recall and perhaps also higher JOLs? We examined this question for each study.

3.1. Study 1

All children in each grade were divided at the median of mean number of TTAs (2.57 for 2nd graders and 2.06 for 4th graders) into fast learners and slow learners. For 4th graders, recall was higher for fast learners (91.3%) than for slow learners (73.7%), $t(18) = 3.89$, $p = .001$. The same was true for 2nd graders (the respective values were 87.1% and 72.9%), $t(18) = 4.00$, $p < .001$.

The results for JOLs were in the same direction, but the difference was not significant. For 4th graders JOLs for fast and slow learners averaged 65.4% and 64.0%, respectively, $t(18) = 0.31$, $p = .76$. The respective means for 2nd graders were 69.7% and 60.9%, $t(18) = 1.59$, $p = .13$.

3.2. Study 2

The results for Study 2 were similar. The fast learners (with TTA < 3.9 for 2nd graders and TTA < 3.2 for 4th graders) exhibited better recall than slow learners, 80.0% and 50.4%, respectively, for 4th graders, $t(28) = 5.60$, $p < .0001$, and 55.6% and 39.2%, respectively, for 2nd graders, $t(28) = 2.31$, $p < .05$. Again, the differences in JOLs were in the same direction but the difference was not significant. JOL for fast and slow learners averaged 62.8% and 54.0%, respectively, for 4th graders, $t(28) = 1.80$, $p < .09$, and 54.0% and 51.8%, respectively, for 2nd graders, $t(28) = 0.60$, $p = .55$.

In sum, the results for individual differences for both studies indicated that the inverse relationship between recall and TTA held in a between-person analysis. Children who needed more TTAs to acquire the entire list also exhibited poorer final recall than those who needed fewer TTAs. Thus, the ELER principle holds true in a cross-item analysis as well as in a cross-individual analysis. JOLs, in contrast, did not yield an inverse relationship in the between-person analysis. Fast learners did not report higher JOLs than slow learners.

4. Discussion

Koriat and his associates (Koriat, 2008a; Koriat et al., 2006) examined the relationship between ease of learning and subsequent recall performance among college students. Ease of learning was operationalized in two ways, first, as the amount of time spent studying an item under self-paced instructions, and second, as the number of trials needed to master the item during study. In both cases a correlation was observed between learning and remembering such that recall was worse for items to which relatively *more* study time was allocated and for items that required relatively *many* TTAs. These results support the validity of ease of learning as a predictor of successful recall. In parallel, the results suggest that learners' metacognitive judgments incorporate knowledge about the link between learning and remembering: JOLs were *inversely* related to self-paced study time as well as to the number of TTAs.

Koriat et al. (2009) extended the investigation to children, focusing on self-paced learning. They found that study time was indeed a valid predictor of subsequent recall performance even among 1st- and 2nd-graders, but only for 3rd–6th graders was JOL inversely related to study time. These results suggested an age development in the use of study effort as a basis for JOLs. It should be noted that even 6th graders did not consistently disclose knowledge of the rule that increased study time is associated with better recall, so that reliance on study time in making JOLs was not based on declarative metacognitive knowledge.

In the present study we examined whether 2nd and 4th graders also yield evidence for the ELER principle, such that recall will decrease with TTA, and whether their JOLs disclose knowledge of that principle. Results paralleled in part those obtained for study time. With regard to cue validity, the results of Study 1 clearly indicated that number of TTAs was indeed a reliable predictor of recall success: for both 4th graders and 2nd graders recall likelihood *decreased* substantially with TTA. These results suggest that although the commonly used dropout procedure in learning experiments (Karpicke & Roediger, 2008) ensures perfect recall of each item, the learning history of an item is nevertheless a powerful predictor of the likelihood that the item will be recalled in the future. Thus, even the results for children run counter to the adage "Easy comes, easy goes."

The results for JOLs (Study 1) also evidenced a clear effect of TTA. For both 4th graders and 2nd graders, JOLs decreased reliably with TTA, suggesting that the ELER principle is incorporated into children's metacognitive monitoring of their own competence during learning.

In order to obtain some clues regarding the mechanism underlying the ELER principle in college students, Koriat (2008a) used only unrelated items to examine whether the ELER principle is entirely mediated by inter-item differences in normative difficulty. Results indicated that both recall and JOLs decrease with TTAs. Furthermore, judges presented the same items failed to predict the relative number of TTAs participants had needed to master different items.

In Study 2, we also included only unrelated word pairs. Results for cue validity were similar to those obtained for college students. For both grades, final recall decreased reliably with TTA, and there was little difference between the two grades in this respect. Results for JOLs, in contrast, suggested an age-related development. Whereas for 4th graders JOLs decreased reliably with increasing TTA, 2nd graders yielded no sign of such a decrease. It is notable, however, that children in both grades failed to estimate the relative number of TTAs that they had needed to master each item. The within-person correlation between actual and estimated number of TTAs (EOL ratings) was close to 0 for both grades.

What are the cues that mediate the relationship between JOL and TTA? Unfortunately, neither the present study nor the study with young adults (Koriat, 2008a) provides a clear answer to this question. There are three candidate cues that may allow learners to make recall predictions that are sensitive to differences in TTAs. First, learners may base their JOLs directly on the estimated number of study trials they had needed to master each item. The results so far do not support this possibility because children did not disclose the belief that items that require more TTAs are less likely to be recalled than those that require fewer TTAs. Furthermore, in Study 2, even 4th graders failed to estimate the number of TTAs associated with each item. A second candidate cue is the judged intrinsic difficulty of the item. Learners may apply the rule that items judged to be difficult on a priori grounds (e.g., items with very weak associative link between the cue and the target) are less likely to be recalled. Indeed, normative item difficulty has been found to affect both JOLs and recall (Koriat, 1997) and may mediate the relationship between them and TTA. As noted earlier, however, this account is not consistent with the observation that the ELER principle was found for unrelated pairs.

A third candidate cue is what we may dub "subjective" or idiosyncratic item difficulty. It refers to the ease with which an item is encoded, as revealed ad hoc to the learner when attempting to commit the item to memory. In the study by Koriat et al. (2009), in which JOLs were made immediately after each self-paced study trial, the idiosyncratic experience of ease or effort can exert a direct effect on JOLs. In the present study, in contrast, the mnemonic cues underlying the JOL–TTA relationship are less clearly identifiable. We propose that in this case JOLs are based on the cumulative feedback that learners collect about the ease with which they manage to encode the item during repeated study trials or to retrieve it during repeated test trials. The assumption is that items that require relatively few TTAs induce a strong feeling of mastery, whereas items that require more TTAs are associated with an overall feeling of effort and frustration. This account is consistent with the general idea that JOLs are based on the experienced ease or effort associated with learning (Benjamin & Bjork, 1996; Koriat et al., 2006, 2009). Perhaps then, as children grow older they not only acquire metacognitive knowledge that can help them in making deliberate inferences and in regulating their behavior (Schneider, 2008) but also develop heuristics that affect their metacognitive state.

Previous research on metacognitive development has emphasized the importance of explicit metacognitive knowledge and understanding and the deliberate application of that knowledge in the strategic regulation of cognitive processes (Kuhn, 2000; Schneider & Pressley, 1997). In parallel, however, there are metacognitive processes that operate implicitly, relying on heuristics and cues that shape one's metacognitive state directly and unconsciously. These implicit processes would seem also to undergo some development, by incorporating implicitly acquired knowledge about the structure of the "internal ecology" (Koriat, 2008a). This development may be similar to that underlying spatial perception, in which emerging sensitivity to various depth cues that are used implicitly to shape perceptual experience helps infants perceive the world correctly (Yonas, Elieff, & Arterberry, 2002). In the case of metacognition, it would seem that the increased sensitivity with age to the feedback that one gains from study experience is more detectable when items are not clearly differentiated in terms of commonly shared properties.

It is important to stress the counterintuitive nature of the ELER heuristic, because this heuristic runs counter to the established observations that memory performance improves with repeated practice and with longer study time. This counterintuitive nature brings to the fore two theoretically important features of the underlying process. First, the ELER heuristic derives its diagnostic value from the fact that the number of trials needed for mastering, an item as well as the amount of study time invested in an item under self-paced study, are *data-driven*, determined by the item itself. Hence both self-paced study time, and number of TTAs are diagnostic of the subjective difficulty of the item for the learner and are inversely related to future recall. In contrast, when variations in amount of practice or amount of study time are *goal-driven*, determined deliberately and willfully by the learner, a *positive* rather than a negative correlation should be observed between TTA and study time, on the one hand, and JOL and recall, on the other (Koriat et al., 2006). Such a correlation should also be obtained when number of study trials or amount of study time are experimentally manipulated.

A second feature brought to the fore by the counterintuitive nature of the ELER heuristic is that learners do not seem to be entirely aware of this heuristic. Indeed, the self-reports of 6th graders in Koriat et al.'s (2009) study indicated that they did not explicitly subscribe to the belief that recall should decrease with self-paced study time. In this study most of the children were not cognizant of the ELER principle even when they were questioned immediately after the experiment. Thus, the JOL–TTA relationship reported here is probably not mediated by the application of a metacognitive belief but rests on a heuristic that operates below full consciousness to influence and shape sheer subjective feelings (Koriat, Nussinson, Bless, & Shaked, 2008). Experience-based metacognitive judgments that are based on the feedback from control operations may have important effects on subsequent control processes, possibly more so than those that might be based on declarative metacognitive knowledge (Koriat, Bjork, Sheffer, & Bar, 2004).

In conclusion, the results of the present study are consistent with those of Koriat et al. (2009) in suggesting that metacognitive judgments rely on the feedback from study experience and that the sensitivity to one's own experience increases with age during childhood.

Acknowledgements

This study is part of a project supported by a grant from the German-Israel Foundation for Scientific Research and Development (Grant No. 942-332.4/2006) and by the Max Wertheimer Minerva Center for Cognitive Processes and Human Performance at the University of Haifa. We are grateful to the Chief Scientist of the Israeli Ministry of Education and Culture for approval of the project. We thank Rinat Gil and Dana Klein for their help in all phases of this study. We are grateful to Janet Metcalfe, Roddy Roediger, and two anonymous reviewers for their helpful comments to an earlier draft.

References

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.

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.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance. Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

Brown, A. L. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 103–152). New York: Academic Press.

Brown, A. S. (2008). Tip of the tongue experience. In H. L. Roediger III (Ed.), *Cognitive psychology of memory. Vol. 2 of learning and memory: A comprehensive reference (J. Byrne, Editor)* (pp. 378–385). Oxford: Elsevier.

Brunswik, E. (1956). *Perception and the representative design in psychological experiments*. Berkeley, CA: University of California Press.

Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, *15*, 692–712.

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., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 374–380.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*, 906–911.

Ghetti, S. (2003). Memory for nonoccurrences: The role of metacognition. *Journal of Memory and Language*, *48*, 722–739.
Ghetti, S. (2008). Rejection of false events in childhood: A metamemory account. *Current Directions in Psychological Science*, *17*, 16–20.
Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 22–34.
Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968.
Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*, 1–24.
Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling-of-knowing. *Psychological Review*, *100*, 609–639.
Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370.
Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge, UK: Cambridge University Press.
Koriat, A. (2008a). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, *36*, 416–428.
Koriat, A. (2008b). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 945–959.
Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*, 478–492.
Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643–656.
Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*, 36–69.
Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *A handbook of memory and metamemory* (pp. 117–135). Mahwah, NJ: Erlbaum.
Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The memorizing effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology*, *102*, 265–279.
Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science*, *9*, 178–181.
Lockl, K., & Schneider, W. (2002). Developmental trends in children's feeling-of-knowing judgements. *International Journal of Behavioral Development*, *26*, 327–333.
Lucangeli, D., Galderisi, D., & Cornoldi, C. (1995). Specific and general transfer effects following metamemory training. *Learning Disabilities Research & Practice*, *10*, 1–21.
Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*, 109–133.
Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 676–686.
Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment Recall And Monitoring (PRAM). *Psychological Methods*, *9*, 53–69.
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.
Schneider, W. (1999). The development of metamemory in children. In D. Gopher & A. Koriat (Eds.), *Attention and performance. XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 487–514). Cambridge, MA: MIT Press.
Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*, *2*, 114–121.
Schneider, W., & Bjorklund, D. F. (1998). Memory. In W. Damon, R. S. Siegler, & D. Kuhn (Eds.), *Handbook of child psychology* (pp. 467–521). New York: Wiley.
Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: Evidence for developmental trends. In J. Dunlosky & R. A. Bjork (Eds.), *A handbook of memory and metamemory* (pp. 391–409). Mahwah, NJ: Erlbaum.
Schneider, W., & Pressley, M. (1997). *Memory development between 2 and 20* (2nd Ed.). Hillsdale, NJ: Erlbaum.
Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage model. *Memory & Cognition*, *33*, 1116–1129.
Yonas, A., Elieff, C. A., & Arterberry, M. E. (2002). Emergence of sensitivity to pictorial depth cues: Charting development in individual infants. *Infants Behavior & Development*, *25*, 495–514.