Journal of Experimental Child Psychology 102 (2009) 265–279

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

The memorizing effort heuristic in judgments of learning: A developmental perspective

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

[a] Department of Psychology, University of Haifa, Mount Carmel, 31905 Haifa, Israel
[b] Department of Psychology, University of Bamberg, D-96045 Bamberg, Germany
[c] Department of Psychology, University of Würzburg, D-97074 Würzburg, Germany

ARTICLE INFO

Article history:
Received 28 October 2007
Revised 7 August 2008
Available online 11 December 2008

Keywords:
Metamemory
Metacognitive development
Monitoring
Control processes
Learning
Self-regulation

ABSTRACT

Recent work on adult metacognition indicates that although metacognitive monitoring often guides control operations, sometimes it follows control operations and is based on the feedback from them. Consistent with this view, in self-paced learning, judgments of learning (JOLs) made at the end of each study trial *decreased* with the amount of time spent studying the item, suggesting that JOLs are based on the memorizing effort heuristic that easily learned items are more likely to be remembered. Study 1 extended investigation to primary school children. Whereas for third to sixth graders (9- to 12-year-olds) JOLs decreased with increasing study time (ST), no such relationship was found for first and second graders (7- and 8-year-olds). For both age groups, however, recall decreased with ST, supporting the validity of the memorizing effort heuristic. Self-reports (Study 2) disclosed the belief that recall should tend to increase with ST. The results bring to the fore the importance of mnemonic cues that shape metacognitive feelings even among primary school children. These cues lie in the very feedback that learners gain on-line from task performance rather than in metacognitive knowledge, and their use may also contribute to increased monitoring accuracy with age.

Introduction

There has been a great deal of work during recent years on the metacognitive processes underlying learning (see Koriat, 2007, for a review). Research with children and adults has explored the possible determinants of learners' feelings of mastery and competence during study and the processes that

* Corresponding author. Fax: +972 4 8249431.
E-mail address: akoriat@research.haifa.ac.il (A. Koriat).

0022-0965/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jecp.2008.10.005

contribute to the accuracy and inaccuracy of these feelings (see Dunning, Johnson, Ehringer, & Kruger, 2003; Koriat, 1997; Schwartz, 1994). The assumption underlying such research is that the monitoring of one's own learning and the regulation of learning strategies may have pronounced effects on memory performance (Benjamin & Bjork, 1996; Brown, 1975; Flavell, 1979; Schneider, 1999; Schneider & Pressley, 1997).

In a typical experiment on metacognitive monitoring during learning, participants study a list of paired associates and at the end of each study trial they make a judgment of learning (JOL) reflecting the likelihood that they will recall each target word at test when probed with its cue word. Many studies that have used this procedure with young adults indicated that learners' JOLs are moderately accurate in predicting which items will be recalled and which ones will not (Dunlosky & Nelson, 1994). Furthermore, under self-paced conditions, learners typically allocate more study time (ST) to items associated with low JOLs than to items associated with high JOLs (e.g., Son & Metcalfe, 2000), and when given the option to restudy some of the items, they generally choose to restudy low-JOL items (Metcalfe & Finn, 2008; Thiede & Dunlosky, 1999). These results suggest the operation of an adaptive process in which learners monitor on-line the degree of learning of each item and regulate the allocation of learning resources in accordance with the monitoring output (Dunlosky & Hertzog, 1998). Indeed, manipulations that improve learners' monitoring accuracy result in a more effective regulation of study and, in turn, in overall better test performance (Thiede, Anderson, & Therriault, 2003).

Developmental studies suggest an age-related improvement in both monitoring and self-regulation during learning. With regard to monitoring, first to fourth graders were found to assign higher JOLs to judged easy paired associates than to judged difficult paired associates (Koriat & Shitzer-Reichert, 2002; Lockl & Schneider, 2003). Even kindergartners' JOLs were found to have some degree of accuracy in predicting subsequent recall (Schneider, Visé, Lockl, & Nelson, 2000), but monitoring accuracy tends to increase with age (Koriat & Shitzer-Reichert, 2002).

With regard to self-regulation, Masur, McIntyre, and Flavell (1973) found that third graders, but not first graders, selected to restudy items that were not recalled correctly on a previous study test trial. Dufresne and Kobasigawa (1989), who asked children to study either easy or hard paired associates until they were sure they could remember all pairs perfectly, found that fifth and seventh graders spent more time studying the hard pairs than they did studying the easy ones, whereas first and third graders spent roughly the same amount of time on both types of pairs. A similar age-related difference was observed by Lockl and Schneider (2002b), Lockl and Schneider (2004) in comparing first and third graders. An interesting observation in Dufresne and Kobasigawa's (1989) study was that most of the first and third graders were able to distinguish between the hard and easy pairs, with many of them even suggesting that the harder pairs should be studied for a longer time. These observations were taken to suggest that the main developmental change occurs in the application of metacognitive knowledge to strategic self-regulation.

A similar conclusion was reached by Lockl and Schneider (2003). For both first and third graders, JOLs made during the first study trial predicted the amount of time that the children invested in each item in a subsequent self-paced study trial, suggesting that even first graders made use of their JOLs in regulating ST. The JOL–ST correlation, however, was significantly stronger for third graders, suggesting that the critical development lies in the ability of children to translate monitoring into adequate self-regulation.

Overall, the developmental results just reviewed, as well as those obtained with adults, are in line with the "monitoring affects control" hypothesis (Nelson & Narens, 1988), which is claimed to apply to a variety of metacognitive judgments (Nelson & Narens, 1990; Son & Schwartz, 2002). For example, the tendency of learners to invest more ST in the more difficult items has been seen to disclose strategic self-regulation. Learners monitor the difficulty of different items in advance of learning and use the products of their monitoring as a basis for allocating ST to different items. Alternatively, according to the discrepancy reduction model (Dunlosky & Hertzog, 1998), learners continuously monitor the on-line increase in encoding strength that occurs as more time is spent studying an item and cease study when a desired level of strength has been reached. To achieve this desired level, more ST should be invested in difficult items than in easy ones.

A different view regarding the monitoring–control relationship was advanced by Koriat, Ma'ayan, and Nussinson (2006). Unlike the general assumption that control operations are based on monitoring, they proposed that sometimes monitoring itself is based on the internal feedback from control operations so that it follows control rather than necessarily preceding it. Thus, it was argued that in the case of self-paced learning, the allocation of ST to different items is generally *data driven*—determined by the studied items or, more precisely, by the learner–item interaction. Learners spend as much ST as they believe is needed to commit an item to memory. Perhaps an analogy can help. Studying an item is like lifting a suitcase in that if one believes that a suitcase is lighter than it actually is, this would not change the amount of effort that one must invest in lifting the suitcase. The effort invested is determined almost entirely by the suitcase itself; therefore, the amount of effort invested in lifting the suitcase provides feedback that is informative about the weight of the suitcase (as argued by Kahneman, 1973, with regard to the effort invested in easy and difficult tasks). Thus, the ST devoted to the study of a particular paired associate depends on the idiosyncratic interaction between the learner and the item. It is determined ad hoc by the ease with which the learner succeeds in forming or discovering a link between the cue and the target; therefore, it conveys to the learner the ease or difficulty with which the item is committed to memory. Learners are assumed to base their JOLs on ST (or study effort) under the memorizing effort heuristic that easily encoded items are more likely to be remembered than are items that require greater effort to study. This "control affects monitoring" hypothesis predicts that under self-paced instructions, JOLs made at the end of each study trial should *decrease* with increasing ST. Indeed, several experiments with college students yielded an *inverse* relationship between ST and JOLs, suggesting reliance on the memorizing effort heuristic in making JOLs. Furthermore, this heuristic was found to be valid in that recall was also inversely related to ST (Koriat, 2008; Koriat, Ma'ayan, & Nussinson, 2006).

The findings that JOLs and recall decrease with ST run counter to the common belief that study effort should enhance memory. In fact, even 4-year-olds have been found to hold the belief that increased effort will lead to increased recall (O'Sullivan, 1993; Wellman, Collins, & Glieberman, 1981). Indeed, Koriat and colleagues (2006) also postulated that both JOLs and recall should *increase* with ST, but this should occur when effort is goal driven rather than data driven. Consistent with this assumption, they found that when different incentives were attached to the recall of different items, young adults invested relatively more ST in the high-incentive items and, in parallel, predicted better recall of these items, thereby yielding a *positive* relationship between ST and JOLs. For each incentive level, however, JOLs and recall *decreased* with the amount of time spent studying an item (Koriat, Ma'ayan, & Nussinson, 2006, Experiment 5), consistent with data-driven regulation. The occurrence of a positive and a negative ST–JOL relationship within the same situation underscores the importance of the distinction between goal-driven and data-driven regulation. This distinction bears some similarity to the distinction between effortful (or proactive) control and reactive control (see Derryberry & Rothbart, 1997; Eisenberg & Morris, 2002). Thus, the common belief that effort improves recall seems to hold true only for goal-driven proactive regulation.

It should be noted that a similar pattern of results was obtained for the relationship between confidence judgments and response latency (Koriat et al., 2006, Experiment 7). When participants were given several psychometric problems to solve, they invested more time in the items that were associated with a higher incentive than in those that were associated with a lower incentive and, in parallel, reported higher confidence in their solutions for the former items than for the latter ones. This positive correlation is consistent with goal-driven regulation. However, for each level of incentive, confidence decreased with solution time, suggesting that it was based on the feedback from task performance (data-driven regulation).

This study focuses on data-driven regulation during self-paced learning. It addresses the following question: At what age do children acquire the heuristic that the more time they devote to the study of an item, the *less* likely they are to recall that item? The results obtained with adults invite a more complex conceptualization of children's monitoring and control processes than is currently available in the literature on children's metacognition. That literature has placed a greater emphasis on children's stated metacognitive beliefs, assuming that such beliefs generally mediate strategic regulation. Research on adult metacognition, in contrast, has emphasized the contribution of mnemonic cues and heuristics that operate below full consciousness to yield a sheer subjective feeling (see Kelley & Jacoby, 1996; Koriat, Nussinson, Bless, & Shaked, 2008). Memorizing effort, as indexed by ST, is one such mnemonic cue that may be used by learners without their being aware of the rule that easily learned items are better remembered than those that require more effort to learn (Koriat, 2008).

In Study 1, children of different ages studied a list of paired associates under self-paced instructions, making JOLs at the end of each study trial. Their cued recall performance was then tested. We examined the developmental changes in both cue utilization and cue validity. *Cue utilization* is reflected in the ST–JOL correlation. This correlation should be negative to the extent that children rely on the memorizing effort heuristic in making JOLs. *Cue validity,* in turn, is reflected in the ST–recall correlation. This correlation should also be negative if the easily learned–easily remembered (ELER) heuristic (Koriat, 2008) has some degree of validity. Of course, to the extent that increased ST is diagnostic of poorer memory, reliance on the ELER heuristic should contribute to monitoring accuracy—the accuracy of JOLs in predicting recall performance.

In Study 2, we examined sixth graders' beliefs regarding the relationship between ST investment and recall success. The results can help to indicate whether indeed children are not aware of the ELER heuristic that may underlie their on-line JOLs.

Study 1

In designing this study, we initially hypothesized that the critical development in the reliance on memorizing effort as a cue for JOLs should occur between the third and fifth grades. Accordingly, the study was designed to compare second and third graders with fifth and sixth graders. Preliminary results, however, suggested that the critical development occurs at an earlier age than we anticipated. Thus, during a second phase of the study, a new group of first graders was added toward the end of the school year so that the study ended up including first, second, third, fifth, and sixth graders, with a prediction that the expected change would occur between the second and third grades. Thus, in what follows, we refer to first and second graders as the *younger* group and to third to sixth graders as the *older* group.

Method

Participants

A total of 100 Israeli children from elementary schools in Israel, 20 in each grade, participated in the study. They were mostly of middle-class and upper middle-class socioeconomic backgrounds. Their mean ages were 7.0 years for the first graders, 7.9 years for the second graders, 8.9 years for the third graders, 10.9 years for the fifth graders, and 12.0 years for the sixth graders.

Materials

The items were 24 pairs of Hebrew words that had been used in previous research (Koriat & Shitzer-Reichert, 2002). Of these 24 pairs, 12 represented easy pairs; in each of them, the words were semantically or associatively related (e.g., *chicken–egg, king–crown*). In the remaining 12 hard pairs, the two members of each pair were unrelated (e.g., *stove–flag, cake–rug*). The pairs were selected on the basis of a norming study in which memorability ratings of the pairs were collected. In that study, 30 second graders and 30 fourth graders rated 50 Hebrew pairs in terms of memorability. They were asked to imagine that 100 children had studied the list of 50 pairs and to estimate for each pair how many of the children are likely to recall the target word in response to the cue word at test. The 24 pairs were selected to represent different degrees of judged memorability, but each of the related pairs received a higher mean memorability rating than each of the unrelated pairs.

Procedure

The consent of the parents and the school was obtained before beginning the study. Children were tested individually in a quiet room in the school using a PC-compatible laptop computer. They were told that they would need to 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. They were instructed to study each pair for as long as they needed and to click with the mouse on a box labeled "continue" when they were through studying the pair. At the end of each study trial, they were asked to assess the likelihood of recalling the response word during the test phase in response to the cue. The elicitation of JOLs capitalized on the cold–hot game familiar to children using a thermometer procedure (see Koriat & Shitzer-Reichert, 2002). The rules of the cold–hot game were explained. Children were required to make their rating by sliding a pointer on a colored scale (in the shape of a horizontal thermometer) using the mouse. The pointer was initially positioned in the middle of the scale (colored white), and children could slide it toward one end of the scale (deep blue for "very cold") or toward the other end (deep red for "very hot").

The 24 word pairs appeared in one of four orders, counterbalanced across participants. Two practice trials were used to illustrate the procedure. On each trial, the children pressed a button when ready to study a word pair. The two words then appeared side by side until the children clicked on the "continue" box, at which time the word pair disappeared. Then the question "How sure are you that you will recall the second word later when you see the first word?" appeared on the screen with the thermometer scale underneath. After marking the JOL rating, the children clicked on a box labeled "next pair" to initiate the next trial.

When the study phase was completed, a 1-min filler task (making a free line drawing) was used. During the test phase that followed, each of the cue words was presented in turn, in a random order, and the recalled responses were spoken orally by the children and then entered by the experimenter on a keyboard. 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. Immediately after these items, as well as at the end of the testing phase, the children were asked to make an aggregate JOL expressing their estimate of the number of items that would be/were recalled correctly.

Measures

The main measures used in the analyses are ST, JOL, and recall. ST was used as an index of memorizing effort. It was defined as the time (in seconds) from the presentation of the item until the children clicked on the "continue" box. JOLs were used to index monitoring—the learners' predictions of recall likelihood. They were derived from the position of the pointer on the cold–hot scale. This position was transformed into a JOL percentage score so that the deep blue end (very cold) was defined as 0% and the deep red end (very hot) was defined as 100%. Recall performance was coded for each item as 100 (correct) or 0 (omission or commission error). A word that was very closely related to the target word (e.g., *shoes* instead of *shoe*) was scored as a correct response.

Results and discussion

Table 1 presents the means of ST, JOL, and recall for each grade for the hard (unrelated) and easy (related) paired associates. We examine the results for each of the variables in turn.

Study time

A two-way analysis of variance (ANOVA), Grade (5) $\times$ Item Difficulty (2: easy vs. hard), was carried out on ST, with grade as a between-participant factor and item difficulty as a within-participant factor. The results yielded $F(4, 95) = 3.42$, $MSE = 41.67$, $p < .05$, for grade, $F(1, 95) = 27.88$, $MSE = 5.47$, $p < .0001$, for item difficulty, and $F(4, 95) = 1.79$, $MSE = 5.47$, $p < .14$, for the interaction. The effects of grade were not systematic. It might have been expected that the younger children would compen-

Table 1
Mean study time, JOL, and recall for hard (unrelated) and easy (related) pairs for each of the grades

Grade	Study time (s)		JOL		Recall (%)	
	Hard	Easy	Hard	Easy	Hard	Easy
1	10.9	9.4	59.7	77.0	10.8	66.3
2	7.8	7.2	63.1	70.6	15.8	68.8
3	7.3	5.7	58.2	79.8	22.1	81.7
5	7.4	5.7	56.1	74.2	27.5	85.4
6	11.9	8.6	52.3	70.4	31.7	85.8

sate for their poorer memory by investing more ST than the older children, but that was not the case. So far as the effects of item difficulty are concerned, more ST was devoted to the hard items (9.1 s) than to the easy ones (7.3 s). This effect, however, was more pronounced for the older children (third to sixth graders) than for the younger children (see Table 1). Planned contrasts between the two age groups on the difference in ST between easy and hard items (the difference was 1.0 s for the younger group and 2.2 s for the older group) yielded $F(1, 95) = 3.08$, $MSE = 10.95$, $p < .09$. In fact, simple t tests indicated significant differences between the hard and easy pairs ($p < .01$) for each of the higher grades but not for either of the lower grades. A similar age-related increase in the sensitivity of ST to intrinsic item difficulty was observed in previous studies (Dufresne & Kobasigawa, 1989; Lockl & Schneider, 2004).

The failure of the younger participants to devote more ST to the harder items contrasts with the observation that even second graders assigned markedly lower memorability ratings (see Method section) to the unrelated pairs (56.6) than to the related pairs (81.5), $t(22) = 19.19$, $p < .0001$. As noted earlier, Dufresne and Kobasigawa (1989) reported a similar observation. The older children in their study, but not the younger children, spent more time studying related paired associates than they did studying unrelated paired associates despite the fact that most of the younger children judged the related pairs to be easier to learn. The authors interpreted these results as indicating that, unlike older children, younger children fail to make use of their metacognitive monitoring in regulating ST between items. However, if the regulation of ST is data driven rather than based on metacognitive knowledge, the age-related differences in ST regulation would seem to carry different implications (discussed later in Discussion section).

Judgment of learning

A similar two-way ANOVA on JOLs yielded $F < 1$ for grade, $F(1, 95) = 129.71$, $MSE = 104.99$, $p < .0001$, for item difficulty, and $F(4, 95) = 2.70$, $MSE = 104.99$, $p < .05$, for the interaction. JOLs were higher for the easy (related) pairs (74.4%) than for the hard (unrelated) pairs (57.9%), and the difference was significant for each of the grades (see also Koriat & Shitzer-Reichert, 2002; Lockl & Schneider, 2003). The interaction reflects the observation that the effect of item difficulty was smaller for the younger children than for the older children. Planned contrasts between the two age groups on the difference in JOL between easy and hard items (12.3 for the younger children and 19.2 for the older children) yielded $F(1, 95) = 5.30$, $MSE = 209.99$, $p < .0001$.

Recall

As would be expected, recall yielded a monotonic increase with age and was better for the related pairs than for the unrelated pairs (Table 1). Thus, a Grade (5) $\times$ Item Difficulty (2) ANOVA yielded $F(4, 95) = 6.32$, $MSE = 493.42$, $p < .001$, for grade, and $F(1, 95) = 652.38$, $MSE = 240.35$, $p < .0001$, for item difficulty. The interaction was not significant, $F < 1$. Recall averaged 72.4 and 21.5% for the easy and hard items, respectively, and the difference was highly significant ($p < .0001$) for each of the grades.

It is instructive to compare the results for recall with those for JOLs. First, it can be seen (Table 1) that JOLs did not exhibit an increase with age as was found for recall. Thus, a two-way ANOVA, Grade (5) $\times$ Measure (2: JOL vs. recall), yielded $F(4, 95) = 7.41$, $MSE = 167.38$, $p < .0001$, for the interaction. To illustrate, whereas recall averaged 39.0% for the younger children and 52.3% for the older children, the respective figures for JOL were 67.6 and 65.2%. These results suggest that the younger children overestimated their memory performance. This conclusion was also supported by the results for aggregate JOLs. The younger children predicted that they would recall 59% of the items and estimated that they had recalled correctly 54% of them. The older children, in contrast, were better calibrated, with the respective figures being 56 and 54%. Thus, the aggregate JOLs also failed to yield an age increase as was observed for recall.

Second, the effects of item difficulty were considerably smaller for JOLs than for recall because of a tendency to overestimate the recall of the hard unrelated pairs. This pattern was observed for each of the five grades. As a consequence, children's JOLs underestimated the effects of intrinsic item difficulty (relatedness) on recall. Indeed, a three-way ANOVA, Grade (5) $\times$ Item Difficulty (2) $\times$ Measure (2), yielded $F(4, 95) = 282.33$, $MSE = 138.13$, $p < .0001$, for the Measure $\times$ Item Difficulty interaction (but $F < 1$ for the triple interaction). Whereas JOLs for hard and easy pairs averaged 57.9 and 74.4%, respectively, across all participants, the corresponding values for recall were 21.6 and 77.6%. Note that in previous studies with young adults, no such discrepancy was observed. In fact, the effects of item difficulty for JOLs were very similar to those observed for recall (for immediate testing) (e.g., Koriat, 1997; Koriat, Bjork, Sheffer, & Bar, 2004).

Cue utilization: Memorizing effort as a cue for JOLs

We now turn to the hypotheses of the study concerning the psychological significance of ST as a subjective index of memorizing effort.

To examine the extent to which children rely on memorizing effort as a basis for their recall predictions, we used a procedure that discloses the within-participant relationship between ST and JOL (see Koriat et al., 2006). All STs were split at the median for each participant, and mean JOLs for below-median STs (short) and above-median STs (long) were calculated for each participant. As may be expected (Table 1), the majority of items in the long category (61%) were hard (unrelated items)—55% for the younger children and 65% for the older children.

Fig. 1A presents for each grade mean JOLs for below-median STs (short) and above-median STs (long). A two-way ANOVA, ST (2: short vs. long) × Grade (5), on the results yielded $F(1, 95) = 23.68$, $MSE = 59.67$, $p < .0001$, for ST, $F < 1$ for grade, and $F(1, 95) = 2.83$, $MSE = 59.67$, $p < .05$, for the interaction. The interaction reflects the observation that only the older children yielded the expected pattern of JOLs *decreasing* with increasing STs. Planned contrasts between the two age groups on the difference in JOL between short-ST and long-ST items (0.94 for the younger children and 8.20 for the older children) yielded $F(1, 95) = 10.69$, $MSE = 119.34$, $p < .01$. JOLs were significantly higher for short-ST items

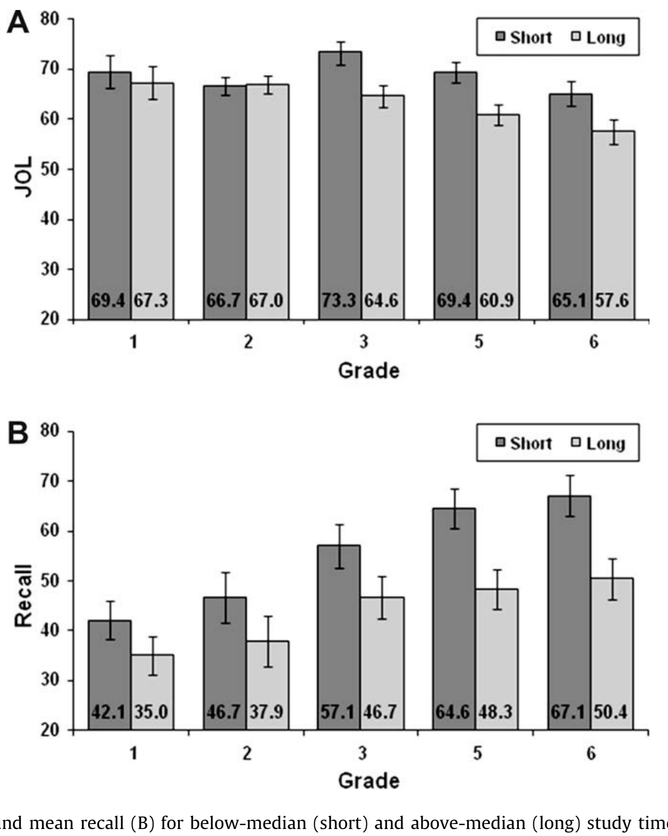

Fig. 1. Mean JOL (A) and mean recall (B) for below-median (short) and above-median (long) study time for each of the five grades.

than for long-ST items for third graders, $t(19) = 3.81$, $p < .005$, for fifth graders, $t(19) = 3.96$, $p < .001$, and for sixth graders, $t(19) = 2.99$, $p < .01$, but not for first graders, $t(19) = 0.66$, *ns*, or for second graders, $t(19) = 0.17$, *ns*. These results suggest that the older children, but not the younger children, relied on the feedback from study experience as a basis for JOLs. It would seem that reliance on the ELER heuristic in making JOLs develops roughly between the second and third grades, that is, around 8.5 years of age.

This conclusion is substantiated by the results for within-participant gamma correlations between ST and JOLs (see Nelson, 1984) that took into account the full range of STs for each participant. These correlations averaged −.07, +.05, −.24, −.29, and −.18 for first, second, third, fifth, and sixth graders, respectively. A one-way ANOVA for the effects of grade on mean gamma correlation yielded $F(4, 95) = 8.13$, $MSE = .045$, $p < .0001$. Planned contrasts indicated that the gamma correlations for the younger children (−.01) differed significantly from those of the older children (−.23), $F(1, 95) = 26.24$, $MSE = 0.04$, $p < .0001$. In fact, for the older children, the correlation was significantly different from 0 for each of the grades: $t(19) = 7.82$, $p < .0001$, for third graders, $t(19) = 6.19$, $p < .0001$, for fifth graders, and $t(19) = 3.50$, $p < .005$, for sixth graders. In contrast, the results for first and second graders were $t(19) = 1.47$, $p < .17$, and $t(19) = 0.81$, $p < .44$, respectively.

It should be acknowledged that the ST–JOL correlation is also consistent with the dominant view in metacognition—that monitoring (JOL) drives control (ST). Because ST is generally strongly correlated with normative item difficulty (Son & Metcalfe, 2000, Table 1), it has been argued that the ST–JOL correlation reflects a process in which learners first judge the relative ease of learning or recalling different items and then deliberately allocate more ST to the judged difficult items than to the judged easy items (Nelson & Leonesio, 1988). This view implies that the typical regulation of ST between items is goal driven, guided by the desire to compensate for differences in a priori item difficulty. But this view cannot accommodate the observation that goal-driven regulation yields diametrically opposed ST–JOL correlations compared with those characteristic of data-driven regulation (Koriat et al., 2006). However, we should admit that the negative ST–JOL correlations, considered alone, are also consistent with the view that it is JOL, based on judged item difficulty, that affects ST.

Cue validity: The validity of ST as a predictor of recall

We now turn to the results on cue validity—the relationship between ST and recall (see Fig. 1B). If children rely on memorizing effort as a cue for JOLs, we may ask how valid memorizing effort is as a predictor of recall. A two-way ANOVA, ST (2) × Grade (5), on recall yielded $F(1, 95) = 34.56$, $MSE = 202.56$, $p < .0001$, for ST, $F(4, 95) = 6.32$, $MSE = 493.42$, $p < .0005$, for grade, and $F < 1$ for the interaction. Thus, recall *decreased* significantly with increasing ST, averaging 52.2% and 41.7% for short- and long-ST items, respectively. It is particularly notable that there was no age difference in the predictive validity of ST.

The validity of the memorizing effort heuristic is also reflected in the ST–recall gamma correlations, which averaged −.13, −.13, −.14, −.23, and −.22 for first, second, third, fifth, and sixth graders, respectively. All correlations were significant ($p < .05$) except the one for the second graders ($p < .11$). A one-way ANOVA revealed no effect of grade on the gamma correlation, $F < 1$. Thus, the more time was invested in studying an item, the less likely it was to be recalled. These results are similar to those obtained for college students (Koriat, 2008; Koriat et al., 2006).

A comparison of the effects of ST on JOL and recall (Fig. 1) suggests that ST was used less as a cue for JOLs than was warranted by the relationship between ST and recall. This was true for children in all grades. A three-way ANOVA, Grade (5) × ST (2) × Measure (2), yielded $F(1, 95) = 12.13$, $MSE = 87.52$, $p < .001$, for the ST × Measure interaction and $F < 1$ for the triple interaction. Across all grades, JOLs for short- and long-ST items averaged 63.5% and 68.8%, respectively. The corresponding means for recall were 43.7% and 55.5%. Thus, children's JOLs did not incorporate the full extent to which recall decreased as a function of ST. It should be noted, however, that the effects of ST were significant for both recall, $t(99) = 5.54$, $p < .0001$, and JOL, $t(99) = 4.70$, $p < .0001$.

Achievement: The accuracy of JOLs in predicting recall

Assuming that participants do rely on memorizing effort in making JOLs and that memorizing effort is diagnostic of recall, we might expect JOLs to exhibit some degree of validity in predicting recall.

Indeed, monitoring accuracy, as indexed by the within-participant JOL–recall gamma correlation (Nelson, 1984), averaged .40, .31, .53, .54, and .64 for first, second, third, fifth, and sixth graders, respectively. Each of these correlations was significantly different from 0 ($p < .005$). Thus, children in all grades were moderately accurate in monitoring their knowledge during study. However, the difference in gamma among the grades was significant, $F(4, 95) = 3.52$, $MSE = 0.095$, $p < .05$, suggesting an age-related increase in the ability to monitor one's own learning (see also Koriat & Shitzer-Reichert, 2002). Indeed a planned contrasts revealed significantly higher correlations for the older children (.57) than for the younger children (.36), $F(1, 95) = 11.56$, $MSE = 0.09$, $p < .01$. We might speculate that this lower accuracy derives in part from the failure of the younger children to use study effort as a cue for JOLs. But the observation that they were nevertheless moderately accurate in monitoring their memory suggests that they were able to rely on other cues for making JOLs such as the judged intrinsic difficulty of the items.

Overall, the results for the older children are consistent with the idea that ST allocation is data driven and that children base their JOLs on ST under the implicit belief that increased ST is diagnostic of lower recall. This belief seems to be counterintuitive in the light of previous results indicating that even very young children expect increased effort to lead to increased recall (O'Sullivan, 1993; Wellman et al., 1981). Study 2 was designed to examine children's stated belief about the relationship between ST and recall. The study was conducted on sixth graders for whom an inverse ST–recall relationship was found in Study 1.

Study 2

To what extent are children aware of the ELER heuristic underlying the inverse relationship between ST and JOL? As noted earlier, it is commonly assumed that experience-based metacognitive judgments are mediated by inferential processes that are not entirely available to consciousness. This is unlike information-based metacognitive judgments, which involve a deliberate and conscious application of a theory or belief (Koriat, 2000; Koriat et al., 2008). In Study 2a, we examined children's beliefs with regard to the ELER principle. Sixth graders were asked to imagine that a friend spends either more ST or less ST on some of the items. They were then asked to judge which of the items he or she is more likely to recall in a subsequent test.

Whereas Study 2a concerned interitem variation in ST, Study 2 b focused on interindividual variation. Sixth graders were asked to imagine that one child spent more time (or effort) studying the list of paired associates than did the other children. They were asked to guess who will exhibit better recall.

Method

Participants

Participants were 90 Israeli sixth graders (45 in Study 2a and 45 in Study 2b). There were 19 boys and 26 girls in Study 2a (mean age of 11.4 years) and 21 boys and 24 girls in Study 2b (mean age of 11.35 years). The children were mostly of middle-class and upper middle-class socioeconomic backgrounds.

Materials, apparatus, and procedure

The studies took place in a classroom. All of the instructions and materials were compiled in a booklet. In Study 2a, participants were asked to imagine that they were watching another child (the gender of the "other" child matched that of the participant) studying eight word pairs under self-paced conditions, spending as much time as he or she needs to master each item. The first question was as follows: "You see that the boy/girl studies some word pairs for a very short time and then moves on to the next item. Do you think that he/she will remember those items better or worse than the other items?" The second question was as follows: "You see that there are pairs that the boy/girl tries particularly hard studying. He/She mumbles, frowns, and looks upward. Do you think that he/she will recall these items better or worse than the other items?" After these two questions, eight word pairs—four related and four unrelated—appeared randomly ordered. The instructions were as follows: "Suppose that the child recalled only four pairs. Mark those that you think he/she recalled."

The procedure for Study 2b was similar except for the instructions for the first two questions. Participants were asked to imagine that they were watching a group of children, with each child seated in front of a computer. The instructions for the first question were as follows: "You see that one boy/girl studies each word pair for a very short time and then moves on to the next item. Do you think that he/she will remember the word pairs better or worse than the other children?" The second question was as follows: "In contrast, you see another child who makes a special effort to study the items. He/She mumbles, frowns, and looks upward. Do you think that the boy/girl will remember the word pairs better or worse than the other children?"

Results and discussion

In Study 2a, 27 participants (60%) judged that an item that was studied for a short period of time will have a lower chance of being recalled relative to the other items ($p < .12$ by a binomial test) and 30 participants (67%) judged that an item in which a great deal of effort was invested will be more likely to be recalled relative to the other items ($p < .05$ by a binomial test). Thus, if anything, participants expected a *positive* within-participant correlation between ST and recall. It should be noted that in Study 1 the JOL–ST correlation was *negative* for 16 of the 20 sixth graders ($p < .01$ by a binomial test).

The results for Study 2b were similar with regard to the questions on ST allocation. In this study, 39 participants (87%) judged that the child who spends only a little amount of ST will exhibit lower recall than the other children ($p < .0001$ by a binomial test) and 38 participants (84%) judged that the child who invested a great deal of effort will exhibit better recall than the other children ($p < .0001$ by a binomial test). Thus, clearly, so far as differences between individuals are concerned, participants expected a positive correlation between mean ST and recall.

The responses to the last question were very clear. As might be expected, nearly all participants in both Study 2a and Study 2b chose the four related pairs as those that will be recalled. Thus, intrinsic difficulty was perceived as the decisive determinant of recall.

In sum, the self-report results obtained in Study 2a generally contrast with those found in Study 1 for the actual within-participant relationship between self-paced ST and JOLs. Whereas in Study 1 JOLs for most sixth graders exhibited a negative correlation with ST, here the explicit questions disclosed, if anything, the opposite belief that increased ST is associated with better recall. It would seem that learners are not aware of the ELER heuristic that seems to underlie the ST–JOL correlation.

Study 2b indicated a stronger positive association still between ST and recall with regard to individual differences. In this case, the children's belief accords with what has been observed by Koriat (2008) for college students. He found that those who spent more ST on average exhibited better recall than those who spent less ST. The same pattern was, in fact, observed in Study 1 for the older children (third to sixth graders). When participants in this group were divided at the median of their mean ST per item, those with above-median ST yielded better recall (62.5%) than those with below-median ST (48.9%), $t(58) = 3.52$, $p < .001$. One explanation for this result is that between-participant differences in self-paced ST reflect differences in motivation. Children with stronger motivation to succeed spend more effort studying each item and consequently exhibit better recall performance (see Koriat, 2008). Therefore, the effects of within-participant variation in ST are in the opposite direction to those of between-participant variation. The difference between these two opposing trends can be seen to illustrate the contrast between data-driven and goal-driven regulation.

It is impressive that sixth graders did discriminate between the situation depicted in Study 2a (within-participant variation) and that depicted in Study 2b (between-participant variation). The association between short ST and poor memory was stronger in Study 2b (87%) than in Study 2a (60%), $\chi^2 = 8.18$, $p < .01$. Similarly, the association between long ST (or more effort) and good memory was stronger in Study 2b (84%) than in Study 2a (67%), $\chi^2 = 3.85$, $p < .05$. Possibly, children assume that differences between learners are more likely to reflect differences in motivation.

General discussion

Relationship between monitoring and control during learning

Underlying much of the work on JOLs in children and adults is the assumption that monitoring affects control (see Nelson & Leonesio, 1988; Son & Schwartz, 2002). It is this assumption that has provided the motivation for investigating on-line monitoring during learning, as expressed by Schneider and colleagues (2000): "The self-monitoring that occurs during learning has a guiding role in the self-paced acquisition of information. In particular, the accuracy of judgments of learning (JOLs) is critical because if the JOLs are inaccurate, the allocation of subsequent study time will be less than optimal" (p. 117).

The recent work by Koriat and colleagues (2006) (see also Koriat, 2008) acknowledges the potential effects of monitoring on control but adds the possibility that monitoring itself may depend on the feedback from control operations. It was argued that in attempting to commit an item to memory, learners do not monitor its intrinsic difficulty and decide in advance how much time to devote to its study. Rather, the amount of time invested is typically data driven. It is determined ad hoc by the item–learner interaction, for example, by the success with which the learner discovers or generates an association between the cue word and target word. This association may be idiosyncratic or commonly shared. Learners then use ST (or study effort) as a cue for JOLs under the heuristic that items requiring longer STs are less likely to be recalled than those requiring shorter STs. In a sense, learners use ST as an internal index of the subjective difficulty (or encoding fluency [see Benjamin & Bjork, 1996]) of the item. This view implies that the monitoring of one's own knowledge is based on the feedback that one gains from attempting to commit an item to memory. In a similar manner, when attempting to solve psychometric problems, participants do not decide beforehand how much time to devote to each problem. Rather, they typically spend as much time as is needed to settle on a solution and then use the solution time as a cue for confidence in the correctness of the answer (Koriat et al., 2006).

Clearly, the evidence for this view is still weak because it rests on correlational data. In particular, ST and JOL are generally correlated with normative item difficulty as captured, for example, by memorability ratings. Therefore, the results are open to the interpretation that it is judged item difficulty (monitoring) that drives ST allocation (control). We should note, however, that similar results to those reported by Koriat and colleagues (2006) were obtained when only associatively unrelated words that differ minimally in normative associative relatedness were used (Koriat, 2008). In that case, ST allocation seems to reflect idiosyncratic differences in ease of encoding (Koriat, 1997; Koriat, 2008). Also, a recent study (Koriat & Nussinson, 2008) in which experienced effort was experimentally manipulated yielded results that are consistent with our conceptualization. Participants in a mental effort condition were asked to contract their eyebrows toward the middle of the forehead while studying a list of paired associates and making JOLs, whereas participants in a control condition were asked to raise their eyebrows. The mental effort participants made significantly lower JOLs than the control participants, consistent with the idea that JOLs are data driven based on the effort experienced during encoding. However, in another experiment where participants were specifically induced to attribute their facial expressions to goal-driven regulation, the reverse pattern was observed, with the mental effort participants reporting significantly higher JOLs than the control participants. Although these observations are consistent with our model, additional efforts to support the model are desired.

The current study

In this study, we focused on the data-driven regulation of ST during self-paced learning, examining it within a developmental perspective. We found that third to sixth graders exhibited the same pattern of ST–JOL relationship that was observed for college students. That is, the recall predictions of even 9-year-olds disclose the heuristic that the more time they spend studying an item, the *less* likely they are to recall it in the future. Furthermore, the recall performance of these children also supports the validity of this heuristic. Longer self-paced STs were predictive of *poorer* recall performance. The correspondence between cue utilization and cue validity that was observed in third to sixth graders possibly contributed to the accuracy of their JOLs in predicting interitem differences in recall performance.

First and second graders, in contrast, yielded little indication for the memorizing effort heuristic. This is consistent with evidence suggesting that strategic competencies and self-regulatory skills typically do not develop before the elementary school years (cf. Schneider & Pressley, 1997) or even before middle childhood.

Is there a further development in data-driven regulation after the school years? A comparison of the results of the older children in this study with those obtained for adults suggests a positive answer. Using data from the first study–test cycle of Experiment 4 of Koriat and colleagues (2006), the gamma correlation between ST and JOLs for college students was −.48, which was significantly higher than the average correlation obtained for the older children in our study (−.23), $t(78) = 5.24$, $p < .0001$. Thus, it seems that there is further improvement with age in the use of ST or study effort as a basis for JOLs. It is interesting to note that, in contrast, the ST–recall gamma correlation was −.29 for the college students and did not differ significantly from that observed for the older children in our study (−.20), $t(78) = 1.58$, $p < .15$. This result suggests a smaller age-related change in the validity of ST as a predictor of recall than in the use of ST as a cue for recall predictions. Perhaps related to this result is the observation that ST was predictive of memory performance even among first and second graders, although these children did not seem to make use of ST as a cue for JOLs. It seems that the use of ST as a cue for JOLs lags behind the developmental stage at which ST is a valid predictor of recall.

Finally, with regard to the self-report studies, the results of Study 2a suggest that the reliance on ST as a cue for JOLs is not based on the explicit deliberate application of learners' metacognitive knowledge; rather, it is based on a heuristic that is used implicitly as a basis for feelings of mastery. This proposal is consistent with the view advanced by Brown and her associates about the nature of metacognitive knowledge that underlies procedural metamemory (Brown, 1975; Brown, Bransford, Ferrara, & Campione, 1983). Possibly, the ease with which an item is mastered instills a sheer feeling that the item is likely to be recalled in the future.

It is interesting, however, that in Study 2b, which focused on between-participant differences in ST, the expected effects of ST on recall were consistent with what was found for the older children in our study and for college students (Koriat, 2008). It would seem that although participants are not aware of the negative ST–recall heuristic characteristic of data-driven regulation, they are cognizant of the positive ST–recall relation that characterizes goal-driven regulation (see Wellman et al., 1981).

What, then, is the nature of the developmental change that we observed around 9 years of age? Consider the effects of item difficulty on ST and JOLs. Children in all grades gave higher JOLs to the easy (related) pairs than to the hard (unrelated) pairs. However, whereas the older children also invested significantly more time studying the hard pairs, the younger children's STs did not differentiate between the two types of pairs. These results are consistent with previous findings (see Dufresne & Kobasigawa, 1989; Lockl & Schneider, 2002b; Lockl & Schneider, 2004) that were interpreted as indicating that young children fail to translate the output of their monitoring to the regulation of ST. This interpretation assumes that the older children's (and adults') sensitivity of STs to item difficulty derives from a deliberate application of metacognitive knowledge to strategic self-regulation. However, if the regulation of ST is data driven, then the sensitivity of ST to intrinsic item difficulty would seem to imply that there is a developmental change either in monitoring or in the very regulation of ST—not necessarily in the application of metacognitive knowledge to the spontaneous regulation of ST.

According to the former possibility, even among the younger children the data-driven regulation of ST is tuned to the inherent difficulties in the studied materials, but only the older children take advantage of the cues from study effort in making recall predictions. In support of this possibility is the finding that a significant ST–recall correlation was found even for first and second graders, suggesting that the ST invested in the items is responsive to some aspects of processing that are predictive of recall. Thus, the observation that for the younger children ST was correlated with recall but not with JOLs suggests an age-related development in monitoring.

The second possibility is that the main development occurs in the quality of self-regulation itself. This possibility is consistent with the position expressed by Schneider and Lockl (2008). Perhaps the allocation of ST by the older children is better tuned to features of the studied materials (including intrinsic item difficulty) that are critical for learning and remembering. Indeed, examination of the regulation of ST between items indicated that the older children devoted a greater amount of their ST to the unrelated items than did the younger children.

These two possibilities, one stressing monitoring and the other stressing regulation, are not mutually exclusive. Perhaps an analogy with the effects of practice can help to clarify the age differences observed. Koriat and colleagues (2006), who presented the same list of items for four study–test blocks, found two changes across these blocks. First, the negative ST–recall correlation increased with practice, suggesting improved tuning to idiosyncratic aspects of the learner–item interaction that disclose the future memorability of the studied items. Second, the negative ST–JOL correlation increased with practice, suggesting increased reliance on ST as a cue for JOLs. Both of these changes appeared to contribute to increased monitoring accuracy (the JOL–recall correlation) with practice. The same pattern of results was obtained by Koriat (2008) when only unrelated word associates were used.

Perhaps similar changes occur in the course of development. First, the data-driven regulation becomes increasingly better tuned to aspects of the learning process that are valid predictors of recall. Second, as children mature, they tend to rely more heavily on their internal mnemonic cues in monitoring their knowledge. These cues lie in the feedback that they receive on-line from task performance (Kelley & Jacoby, 1998). In a sense, children learn to trust their own experience for better or for worse. Both changes should result in improved monitoring accuracy, as was indeed found.

A question that suggests itself is whether JOLs are of little value if they simply mirror the processes that have already occurred. Clearly, this is not the case. As Koriat and colleagues (2006) stressed, monitoring and control processes are interwoven in the course of learning: ‘‘Monitoring and control functions alternate in a cascaded pattern, with control following along in the wake of monitoring and the feedback from the control operation serving then as the input for later monitoring, and so on" (p. 41) (for examples, see Koriat & Levy-Sadot, 2001; Son & Metcalfe, 2005). Thus, JOLs made in the course of self-paced learning can affect the allocation of ST on a subsequent study block of these items and the selection of items for restudy (see Lockl & Schneider, 2003; Metcalfe & Finn, 2008). This is why the accuracy of JOLs is critical for optimal learning.

We conclude by mentioning one general implication of this study for the investigation of metacognition in children. As we noted earlier, the work on children's metacognitive skills has traditionally placed a greater emphasis on information-based metacognitive judgments than on experience-based ones (see Koriat, 2007; Schneider & Bjorklund, 1998; Schneider & Lockl, 2008). The assumption underlying this emphasis is that what children believe and know about memory is critical for the strategic regulation of learning and remembering. Recent research, however, indicates that even college students hardly apply their knowledge and theories in making JOLs. Two recent studies will suffice to illustrate this point. First, Koriat and colleagues (2004) speculated that if JOLs are based on processing fluency during study, they should be insensitive to the anticipated retention interval because the processing fluency of an item at the time of encoding should not be affected by when testing is expected. Indeed, JOLs were entirely indifferent to the expected retention interval, although actual recall exhibited the typical forgetting function. Second, Kornell and Bjork (2006) observed that recall predictions fail to take into account the effects of learning. Although actual recall exhibited the typical learning curve, increasing with repeated study–test cycles, predicted learning curves hardly increased with the study cycle. Thus, even college students do not spontaneously apply some of the most basic theories about learning and remembering in making recall predictions.

Not surprising, research on procedural metacognition among adults has paid greater attention to experience-based JOLs, which are assumed to rely on mnemonic cues such as the fluency with which information is encoded or retrieved (Kelley & Lindsay, 1993; Koriat & Ma'ayan, 2005). The use of such cues generally occurs below full consciousness and gives rise to a sheer experience of knowing. For example, neither the processes assumed to underlie the allocation of ST under self-paced instructions nor the use of the ELER heuristic as a basis for JOLs is necessarily available to consciousness (see Koriat, 2000). In a similar manner, other cues that derive from on-line processing may produce a sheer metacognitive feeling. In fact, in one of the earliest discussions of metacognition, Brown (1975); (see also Brown et al., 1983) stressed the distinction between the declarative component of metacognition, which is largely statable, and the procedural component, which is not necessarily statable.

Fortunately, some of the recent developmental work on procedural metacognition has paid greater attention to the contribution of internal mnemonic cues to metacognitive judgments. Examples include the study of Koriat and Shitzer-Reichert (2002) on the factors that affect the accuracy of mnemonic-based JOLs and the studies of Roebers, von der Linden, Schneider, and Howie (2007) and Lockl and Schneider (2002a), which provide some insight on the basis for children's feeling-of-knowing (FOK) judgments. The current research joins with these studies in bringing to the fore the importance of subtle mnemonic cues that are used by children and adults in monitoring their learning.

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: Lawrence Erlbaum.

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* (Vol. 10, pp. 103–152). New York: Academic Press.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning remembering and understanding. In J. H. Flavell & E. M. Markham (Eds.). *Handbook of child psychology. Cognitive development* (Vol. 3, pp. 77–166). New York: John Wiley.

Derryberry, D., & Rothbart, M. K. (1997). Reactive and effortful processes in the organization of temperament. *Development and Psychopathology, 9*, 633–652.

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., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker (Ed.), *Metacognition in educational theory and practice* (pp. 249–275). Mahwah, NJ: Lawrence Erlbaum.

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

Dunning, D., Johnson, K., Ehringer, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83–87.

Eisenberg, N., & Morris, A. S. (2002). Children's emotion-related regulation. In R. Kail (Ed.). *Advances in child development and behavior* (Vol. 30, pp. 190–229). Amsterdam: Academic Press.

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

Kahneman, D. (1973). *Attention and effort*. Prentice Hall: Englewood Cliffs, NJ.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35*, 157–175.

Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica, 98*, 127–140.

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. (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. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9*, 149–171.

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. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition, 36*, 416–428.

Koriat, A., Nussinson, R., (2008). Attributing study effort to data-driven and goal-driven effects: Implications for metacognitive judgments. Unpublished manuscript, University of Haifa.

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., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and the accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 34–53.

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., 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: Lawrence Erlbaum.

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

Kornell, N., & Bjork, R. A. (2006, November). Objective and subjective learning curves. Talk presented at the 47th annual meeting of the Psychonomic Society, Houston, TX.

Lockl, K., & Schneider, W. (2002a). Developmental trends in children's feeling-of-knowing judgements. *International Journal of Behavioral Development, 26*, 327–333.

Lockl, K., & Schneider, W. (2002b). Zur Entwicklung des selbstregulierten Lernens im Grundschulalter: Zusammenhänge zwischen Aufgabenschwierigkeit und Lernzeiteinteilung [The development of self-regulated learning in elementary school children: Associations between task difficulty and allocation of study time]. *Psychologie in Erziehung und Unterricht, 49*, 3–16.

Lockl, K., & Schneider, W. (2003). Metakognitive Überwachungs und Kontrollprozesse bei der Lernzeiteinteilung von Kindern [Monitoring and control processes in children's allocation of study time]. *Zeitschrift für Pädagogische Psychologie, 17*, 173–183.

Lockl, K., & Schneider, W. (2004). The effects of incentives and instructions on children's allocation of study time. *European Journal of Developmental Psychology, 1*, 153–169.
Masur, E. F., McIntyre, C. W., & Flavell, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrial free recall task. *Journal of Experimental Child Psychology, 15*, 237–246.
Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*, 174–179.
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., & Narens, L. (1988). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.). *The psychology of learning and motivation:* (Vol. 26, pp. 125–173). New York: Academic Press.
Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–173). New York: Academic Press.
O'Sullivan, J. T. (1993). Preschoolers' beliefs about effort, incentives, and recall. *Journal of Experimental Child Psychology, 55*, 396–414.
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., & Bjorklund, D. F. (1998). Memory. In W. Damon, R. S. Siegler, & D. Kuhn (Eds.). *Handbook of child psychology* (Vol. 2, 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: Lawrence Erlbaum.
Schneider, W., & Pressley, M. (1997). *Memory development between 2 and 20* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task. *Cognitive Development, 15*, 115–134.
Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin and Review, 1*, 357–375.
Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 204–221.
Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage model. *Memory & Cognition, 33*, 1116–1129.
Son, L. K., & Schwartz, B. L. (2002). The relation between metacognitive monitoring and control. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 15–38). Cambridge, UK: Cambridge University Press.
Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66–73.
Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024–1037.
Wellman, H. M., Collins, J., & Glieberman, J. (1981). Understanding the combination of memory variables: Developing conceptions of memory limitations. *Child Development, 52*, 1313–1317.