Journal of Experimental Psychology: General
2014, Vol. 143, No. 1, 386–403

0096-3445/14/$12.00 DOI: 10.1037/a0031768

The Effects of Goal-Driven and Data-Driven Regulation on Metacognitive Monitoring During Learning: A Developmental Perspective

Asher Koriat
University of Haifa

Rakefet Ackerman
Technion–Israel Institute of Technology

Shiri Adiv
University of Haifa

Kathrin Lockl
University of Bamberg

Wolfgang Schneider
University of Würzburg

Research in metacognition (Koriat, Ma'ayan, & Nussinson, 2006) suggests bidirectional links between monitoring and control during learning: When self-regulation is goal-driven, monitoring affects control so that increased study time (ST) enhances judgments of learning (JOLs). However, when self-regulation is data-driven, JOLs are based on the feedback from control, and therefore JOLs decrease with ST under the heuristic that ease of encoding is diagnostic of successful recall. Evidence for both types of relationships occurring within the same situation was found for adults. We examined the development of the ability to respond differentially to data-driven and goal-driven variation in ST within the same task. Children in Grades 5 and 6 exhibited a positive ST–JOL relationship for goal-driven regulation and a negative relationship for data-driven regulation but never in the same task. In contrast, the JOLs and recall of 9th graders and college students yielded differential cosensitivity to data-driven and goal-driven variation. The 5th and 6th graders also evidenced an adult-like pattern of JOLs and recall under a partitioning procedure that helped them in factoring the variation in ST due to data-driven and goal-driven variation in ST. The results are discussed in terms of the metacognitive sophistication needed for considering both types of variation simultaneously in making metacognitive judgments.

Keywords: metacognitive development, monitoring and control, judgments of learning, self-regulation, incentives

Historically, there have been two major lines of research on metacognition, one within developmental psychology and the other within experimental memory research. These two lines differ somewhat in their goals and methodological styles (see Koriat & Shitzer-Reichert, 2002). The developmental tradition, which was spurred by Flavell, is predicated on the assumption that metacognitive processes play a critical role in the development of learning and memory (see Flavell, 1979, 1999; Flavell & Wellman, 1977). A major aim in that tradition has been to trace the development of metacognitive knowledge and metacognitive skills and to explore their relations to basic cognitive abilities and performance (see Krebs & Roebers, 2012; Schneider, 2008, 2010; Schneider & Lockl, 2008). Within memory research, the experimental study of adult metacognition has been concerned more narrowly with uncovering the mechanisms underlying the monitoring of one's own knowledge and the consequences of monitoring for the strategic regulation of learning and remembering. Although research on metacognition has proceeded almost independently within the developmental and cognitive–experimental traditions, there has been increasing cross-talk between the two lines of research that enhances mutual fertilization. In particular, experimental paradigms developed in the study of adult metacognition have been extended to children (see Ghetti, 2008; Lockl & Schneider, 2002; Lyons & Ghetti, 2010; Roderer & Roebers, 2008; Roebers, von der Linden, Schneider, & Howie, 2007). The work reported in this study also represents such an extension. It attempts to provide some insight into the dynamics underlying metacognitive monitoring and self-regulation in children

This article was published Online First February 18, 2013.

Asher Koriat, Department of Psychology, University of Haifa, Haifa, Israel; Rakefet Ackerman, Faculty of Industrial Engineering and Management, Technion–Israel Institute of Technology, Haifa, Israel; Shiri Adiv, Department of Psychology, University of Haifa, Haifa, Israel; Kathrin Lockl, Department of Psychology, University of Bamberg, Bamberg, Germany; Wolfgang Schneider, Department of Psychology, University of Würzburg, Würzburg, Germany.

This study is part of a project supported by a grant from the German–Israel Foundation for Scientific Research and Development (Grant 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 for her help in the analyses.

Correspondence concerning this article should be addressed to Asher Koriat, Department of Psychology, University of Haifa, Mount Carmel, Haifa 31905, Israel. E-mail: akoriat@research.haifa.ac.il

and into the cognitive prerequisites for the sophisticated metacognitive processes demonstrated by young adults.

The Monitoring-Control Relationship During Learning

In this study, we examined the development of the relationship between metacognitive monitoring and metacognitive regulation during learning (Nelson & Narens, 1990). Metacognitive monitoring refers to the subjective assessment of one's own cognitive processes and knowledge. For example, learners are assumed to monitor online their mastery of the studied material. Metacognitive control, in turn, refers to the strategic regulation of cognitive operations and resources. For example, learners sometimes choose between different learning strategies, allocate their ST differentially between different parts of the studied material, and select specific parts for restudy. A widely held assumption among students of metacognition is that metacognitive monitoring guides control operations. Therefore, the effective monitoring of one's own learning is critical for the effective self-management of learning (Ackerman & Goldsmith, 2011; Ariel, Dunlosky, & Bailey, 2009; Thiede & Dunlosky, 1999; Tullis & Benjamin, 2011). Indeed, research suggests that when learners are asked to make judgments of learning (JOLs) immediately after studying each item, their JOLs predict which items they will later select for restudy (Kornell & Metcalfe, 2006; Metcalfe & Finn, 2008; Nelson, Dunlosky, Graf, & Narens, 1994). Also manipulations that enhance monitoring accuracy were found to improve the effectiveness of ST allocation between different items as well as overall recall performance (Thiede, Anderson, & Therriault, 2003).

One finding that has been replicated across many studies is that in self-paced learning, participants spend more time studying judged-difficult items than judged-easy items (see Son & Metcalfe, 2000, for a review). This finding has been interpreted typically in terms of the "monitoring-affects-control" hypothesis (Nelson & Leonesio, 1988): Learners aim for a desired level of mastery across items. Therefore, they deliberately allocate more ST to the judged-difficult items in order to compensate for their difficulty (Dunlosky & Hertzog, 1998). This interpretation is consistent with the monitoring→control (MC) model according to which ST allocation is based on the output of monitoring operations and is used by the learner as a strategic tool toward the achievement of specific cognitive goals. The goal-driven regulation of ST can be demonstrated by manipulating the incentives associated with the recall of different items in a list. The results of previous studies indicated that participants allocate more ST to high-incentive than to low-incentive items (Ariel et al., 2009; Dunlosky & Thiede, 1998), and their JOLs increase accordingly with increased ST (Koriat et al., 2006).

Several observations, however, suggest that metacognitive monitoring also may be based on the feedback from control operations so that monitoring actually follows, rather than precedes, control operations (see Kelley & Jacoby, 1998). Koriat (1993), for example, suggested that feelings of knowing (FOK) are based on the feedback from the search for the illusive memory target—the number of partial clues that come to mind and the ease with which they come to mind. The implication is that by searching for a memory target, participants "know" whether an unrecallable item is available in memory. In like manner, Koriat et al. (2006) proposed that it is by attempting to commit an item to memory that learners judge whether they would be likely to recall it in the future. They argued that in self-paced learning, ST allocation is generally data-driven rather than goal-driven: Learners determine the ST allocation on the basis of the item itself. Although ST is highly correlated with indexes of normative item difficulty (e.g., degree of associative strength between the members of the pair, see Mueller, Tauber, & Dunlosky, in press), results suggest that ST allocation is idiosyncratic, reflecting the item–learner interaction (Koriat, 1997, 2008). Thus, learners spend as much time and effort as the particular item "calls for" in a bottom–up fashion. Their JOL is then based retrospectively on the memorizing effort heuristic according to which the more effort needed to study an item, the less likely it is to be recalled. Thus, it is by spending a great deal of effort attempting to commit an item to memory that a learner realizes that the item is "difficult" and is less likely to be recalled in the future.

The data-driven regulation brings to the fore the monitoring function of ST and implies a control→monitoring (CM) model in which the output from metacognitive control serves to inform metacognitive monitoring. It should be stressed that in data-driven regulation, JOLs are assumed to rest not on ST as such, but on subjective effort. However, ST is a good indicator of data-driven study effort (see Koriat et al., 2006).

The contrast between the MC and CM models is reminiscent of the issue raised by William James (1884): Do we run away because we are frightened, or are we frightened because we run away? The MC model accords with the view that subjective feelings (e.g., fear) drive behavior (e.g., running away). James's own position—that feelings are based on the feedback from one's own bodily reactions (see Niedenthal, 2007; Strack & Deutsch, 2004; Strack & Neumann, 2000)—is more consistent with the CM model.

The two types of models that we have sketched for self-paced learning are expected to yield diametrically opposed relationships between JOL and ST: JOLs are expected to *increase* with ST when ST is goal-driven, that is, used as a tool in the service of achieving specific motivational objectives. For example, a student may choose to spend more time studying certain parts of the material than others because she finds them interesting. These parts are likely to be associated with higher recall predictions as well as better actual recall than the other parts. In contrast, when ST is data-driven, JOLs are expected to decrease with ST. In that case, study speed reflects encoding fluency, and the items associated with shorter ST would be expected to exhibit higher predicted and actual recall.

Koriat et al. (2006) argued, however, that the two models are not mutually exclusive. Indeed, they found evidence for both types of ST–JOL relationships within the same task. In one of their studies (Experiment 5), college students were awarded different incentives for the successful recall of different items: 1 point for recalling some of the items, and 3 points for recalling the other items. As can be seen in Figure 1, the manipulation of incentive resulted in a *positive* relationship between JOLs and ST: Participants invested more ST in the 3-point items (5.2 s on average per item) than in the 1-point items (4.3 s) and in parallel assigned higher JOLs to the 3-point items (61.4% on average) than to the 1-point items (57.0%). At the same time, however, a *negative* ST–JOL relationship was observed within each incentive level, so that the more ST was invested in an item, the lower was the JOL associated with that

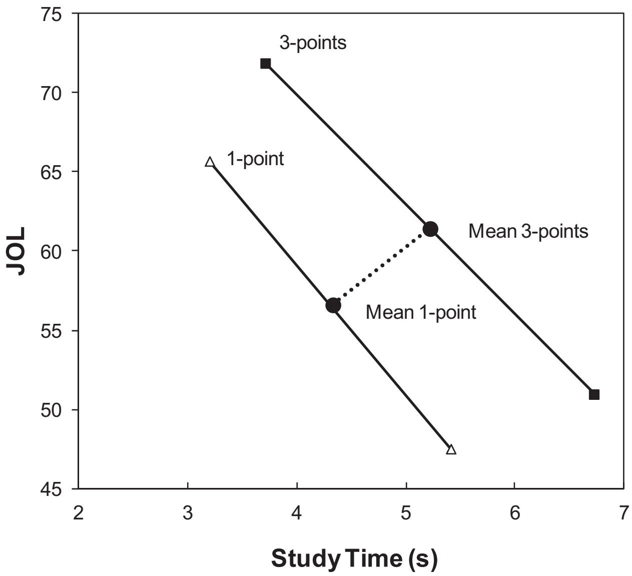

Figure 1. Mean judgment of learning (JOL) for young adults for below-median and above-median study time (in seconds) for each incentive level (solid lines). Plotted also (dotted line) is mean JOL as a function of mean study time for each incentive level (labeled *mean 1 point* and *mean 3 point*). Adapted from "The Intricate Relationships Between Monitoring and Control in Metacognition: Lessons for the Cause-And-Effect Relation Between Subjective Experience and Behavior," by Asher Koriat, Hilit Ma'ayan, and Ravit Nussinson, 2006, *Journal of Experimental Psychology: General, 135*, 36–69, p. 53. Copyright 2006 by American Psychological Association.

item. This relationship is consistent with the CM model suggesting that the allocation of ST between same-incentive items is data-driven. A similar pattern was observed for the relationship between confidence judgments and response latency in a problem-solving task (Koriat et al., 2006, Experiment 7): When participants were given several problems to solve, they invested more time in the problems that were associated with a higher incentive than in those that were associated with a lower incentive. In parallel, they reported higher confidence in the solutions of the former items than in those of the latter items (MC model). However, for each level of incentive, confidence decreased with solution time, suggesting that confidence was based on the feedback from task performance (CM model; see Ackerman & Zalmanov, 2012; Kelley & Lindsay, 1993). These results suggest that the two models considered by William James may not be mutually exclusive.

It is important to note that the pattern of metacognitive monitoring depicted in Figure 1 tends to mirror actual performance. Thus, participants are generally skillful in deriving the implications for recall[1] of variations in ST by attributing these variations to their respective source—data-driven or goal-driven.

The occurrence of a positive and a negative ST–JOL relationship within the same situation implies an attribution process that intervenes between ST regulation and metacognitive monitoring. The component of ST that is attributed by the learner to data-driven effects contributes toward reducing one's JOLs, whereas the component that is attributed to the effects of goal-driven regulation contributes toward enhancing one's JOLs. The reality of this attribution process was brought to the fore by Koriat and Nussinson (2009). They asked learners to adopt a facial expression that creates a feeling of effort and induced them to attribute that effort either to data-driven or to goal-driven regulation. The facial expression of effort was found to decrease JOLs when it was attributed to data-driven regulation but to enhance JOLs when it was attributed to goal-driven regulation.

The present study focused on the ability to consider the metacognitive implications of both data-driven and goal-driven variation within the same task. The simultaneous sensitivity to goal-driven and data-driven regulation is called for in many real-life situations where the amount of effort invested is a joint function of both types of regulation. For example, a learner may invest a great deal of time studying a piece of information partly because of its inherent difficulty and partly because he or she finds it important. What is the developmental trajectory of the ability to consider both data-driven and goal-driven variations in making JOLs, and what are the cognitive prerequisites underlying this ability?

The Development of Metacognitive Judgments

The results presented in Figure 1 suggest that young adults exhibit an impressive ability to respond differentially to goal-driven and data-driven variations in ST. We hypothesized that children's metacognitive ability to take into account the implications of both types of regulation develops at a much later stage than that at which children have been claimed to reach adult-like metacognitive monitoring (e.g., Butterfield, Nelson, & Peck, 1988; Roebers et al., 2007). Some previous studies have confirmed the sensitivity of young children's JOLs to data-driven variations, whereas others have indicated that young children exhibit goal-driven regulation. However, none of these studies has examined the joint impact of both types of regulation on metacognitive judgments.

Koriat, Ackerman, Lockl, and Schneider (2009b) reported a developmental increase in data-driven regulation. They had children study a list of paired associates and indicate their JOLs following the study of each item. JOLs were found to *decrease* with the amount of time invested in the study of each item for children in third through sixth grades children, but not for those in first and second grade. The within-person ST–JOL correlation was negative and significant for third graders (averaging –.24), fifth graders (–.29), and sixth graders (–.18; fourth graders were not included in that study). Although this pattern suggests that JOLs are responsive to data-driven regulation already at the age of 8 or 9 years, there was some further developmental trend thereafter, as suggested by the higher ST–JOL correlation (−.48) observed for young adults (Koriat et al., 2006). These results (see also Hoffmann-Biencourt, Lockl, Schneider, Ackerman, & Koriat, 2010) suggest an age-related increase from first grade to adulthood in the sensitivity of JOLs to the feedback from data-driven regulation.

With regard to goal-driven regulation, even 4-year-olds have been found to disclose the belief that increased effort leads to increased recall (O'Sullivan, 1993, 1997; Stipek, Roberts, & San-

[1] In Experiment 5 of Koriat et al. (2006), recall did not increase with incentive, but the results of unpublished experiments have generally yielded such increase (see also Figure 6 below).

born, 1984; Wellman, Collins, & Glieberman, 1981). This belief was also confirmed for first, third, and fifth graders (Annevirta & Vauras, 2001; Kreutzer, Leonard, & Flavell, 1975). In the study of Koriat, Ackerman, Lockl, and Schneider (2009b), sixth graders expected recall to increase with ST in a self-paced condition although the results for another group drawn from the same population indicated that JOLs (and recall) actually decreased with ST. This result suggests that learners may not be aware of relying on the memorizing effort heuristic that information requiring more time to study is less likely to be recalled (in contrast to the adage "easy come, easy go"; see Koriat, 2008; Koriat et al., 2009a).

There is also some evidence of a developmental trend in the ability of children to actually regulate their performance according to particular goals. Fritz, Morris, Nolan, and Singleton (2007) found that preschoolers did not adjust their learning to the promised reward for remembering. However, Castel, Lee, Humphreys, and Moore (2011) found that the recall of children between 6 and 9 years old was sensitive to the point value attached to each studied item. Using a sample with a wider age range of 6 through 18 years old, Hanten and her associates (Hanten et al., 2004, 2007; Hanten, Zhang, & Levin, 2002) attached different values to the memory of different words. A selective learning efficiency score, reflecting the preferential recall of words of higher value, was found to increase monotonically with age (see also Castel et al., 2011). In all of these studies, items were presented at a fixed rate.

For self-paced learning, Kunzinger and Witryol (1984) found that 7-year-old children allocated more rehearsal to high-value words and exhibited better recall of these words than low-value words. Similar results were obtained in other studies (see Schneider & Pressley, 1997). Lockl and Schneider (2004) found that only third graders, not first graders, studied longer when accuracy was emphasized than when speed was emphasized. However, children in both age groups did not vary their study time in accordance with the presence or absence of incentives. We are not aware of previous developmental studies examining the effects of goal-driven investment of effort on JOLs.

In this study, we chose fifth and sixth graders as our target population because children at these grades not only exhibit evidence for data-driven regulation in self-paced learning but also rely on the feedback from that regulation in monitoring their own knowledge during study (Hoffmann-Biencourt et al., 2010; Koriat et al., 2009b). At the same time, they also show evidence for goal-driven regulation, as suggested by the results of Hanten et al. (2007). Also, in terms of Piaget's (1977) stages, these children are at the end of the concrete operations stage (7–11 years old): They should be able to consider two relevant dimensions simultaneously (Demmrich, 2005; Ojose, 2008; Piaget, 1977). We examined the ability of these children to adjust their JOLs to both data-driven and goal-driven regulation within the same task. Because we had suspected that children at this age might not reveal the pattern of differential sensitivity exhibited by young adults, we proceeded step by step, increasing gradually the complexity of the task.

In all of the following experiments, we manipulated goal-driven variation in ST by awarding different incentives to the recall of different paired-associates. The value of each item was announced prior to its presentation. Data-driven variation in ST, in contrast, was produced by using paired associates that differed in the degree of relatedness between the members of the pairs. JOLs were assessed at the end of each self-paced study trial. If children use ST as a tool for regulating learning, then according to the MC model, ST investment should increase with incentive, and JOLs as well as recall should also yield a corresponding increase. At the same time, in line with the CM model, we expected the results for each incentive level to yield the pattern characteristic of data-driven regulation: decreased JOLs with increased ST.

Experiment 1

Although there is evidence that fifth and sixth graders' JOLs exhibit sensitivity to data-driven variation in ST (Koriat et al., 2009b), there has been no evidence that their JOLs during learning are sensitive to goal-driven variations in ST. Our preliminary attempt to demonstrate such goal-driven effects with fifth and sixth graders using the design of Experiment 5 in Koriat et al. (2006) yielded evidence for the effects of data-driven variation on JOLs, but little evidence for the effects of goal-driven variation. Therefore, in Experiment 1, we sought to enhance the effects of goal-driven variation in comparison to those of data-driven variation by slating all unrelated paired associates to one block and all related paired associates to another block. Incentive was manipulated within each block. In Experiment 2, in contrast, unrelated and related pairs were mixed within each block, and incentive was manipulated between blocks. This design was expected to enhance the contribution of data-driven variation relative to that of goal-driven variation.

Method

Participants. Participants were 40 children from primary schools in Israel, mostly of middle-class and upper middle-class socioeconomic background. The sample included 20 fifth graders (mean age 10.4 years) and 20 sixth graders (mean age 11.4 years), with an overall mean age of 10.9 years.

Materials. The items were 24 pairs of Hebrew words that had been used in previous research (Koriat et al., 2009b). For 12 pairs, the members of each pair were associatively related (e.g., *chicken–egg*, *king–crown*), and for 12 pairs, the two words were unrelated (e.g., *stove–flag*, *cake–rug*). The related and unrelated pairs were slated to two separate blocks so that the assignment of the two sets of items to each block was counterbalanced across participants. Four additional pairs were used for practice.

Procedure. The consent of the parents and of the school was obtained before the beginning of the study. Children were tested individually in a quiet room in the school, using a PC-compatible laptop computer. They were told that the experiment concerned the ability of children to allocate learning resources to different topics according to their importance. To illustrate the experimental task, the children were asked to choose two activities out of four (basketball, ceramics, drama, and martial arts) that they would prefer for after-school activities. They were told to imagine that only a few children would be admitted to each activity, depending on their success in studying passages on topics relevant to that activity. Therefore, each child was asked to study two short-text passages, each relevant to one of the activities that the child had chosen, investing more effort in the most preferred topic than in the less preferred topic. Their memory of each passage was then tested with one simple question for each passage.

Following this demonstration, they were told that they would have to study pairs of words appearing on the computer screen so that they would be able to recall the response word when cued with the stimulus word at the test phase. The importance of each item would be indicated by an incentive value, which represents the number of points that they can earn for correct recall of that item and that the number of points awarded would be indicated with either one star or five stars that would be presented before each item (the analogy with the task of studying the two text passages was mentioned). They were told that their task was to study the list so as to earn as many points as they could. They could study each pair as long as they needed but should try to spend as little time as possible in studying the entire list. To encourage them to take into account the incentive associated with each item, the experimenter mentioned that all the five-star items would appear in the test, but only some of the one-star items would be tested (in fact, only the practice items that were associated with one star were not included in the test phase). Two word pairs were used for practice at the beginning of each of the two blocks.

In each study trial, the child clicked a “show pair” box when ready to study a word pair. Either one star or five blinking stars were then presented, and 2 s thereafter, the study pair was added on the screen and remained with the stars on the screen until the child clicked a box labeled “continue” with the mouse. The time elapsed between the two presses constituted the ST measure. Immediately thereafter, the JOL question appeared: “How sure are you that you will recall the second word later when you see the first word?” The measurement of JOLs capitalized on the hot–cold game familiar to children, using a thermometer procedure (see Koriat et al., 2009b, and Koriat & Shitzer-Reichert, 2002, for details). Children made their ratings by sliding a pointer on a colored scale with the mouse. The position on the scale was transformed into a JOL percentage score (0%–100%). After marking the JOL rating, the child clicked a box labeled “next pair” to initiate the next trial. When the first list was over, a note appeared announcing the beginning of the second list. The 24 word pairs appeared in one of four orders, counterbalanced across participants.

At the end of the second study block, a 1-min filler task was administered (making a free-line drawing). In the test phase that followed, all cue words from the two study blocks were presented each in turn in a random order. The recalled response was spoken aloud by the child and was recorded by the experimenter. The children responded at their own pace without time limit. They were encouraged to try to recall the response word, but when unable to produce a response, they could move to the next cue word. The first two cue words were from the practice items.

Results and Discussion

Preliminary analyses yielded no differences between fifth and sixth graders in this or in the following experiments. Therefore, in all of the analyses, the results were pooled across the two grades.

Figure 2 (Panel A) presents the results for JOLs and recall. In the preparation of this figure, all items associated with the same incentive level (1 point or 5 points) were split at the ST median for each participant, and average JOLs and recall for below-median (short) and above-median (long) STs were calculated. These four means per participant served as the basis for the analyses, unless otherwise stated. The mean JOLs and recall are plotted for the short and long ST items for each of the incentive levels as a function of the mean ST for short and long STs. Goal-driven regulation is depicted by the center lines representing mean JOL/recall for 1-point and 5-point items as a function of their respective mean STs.

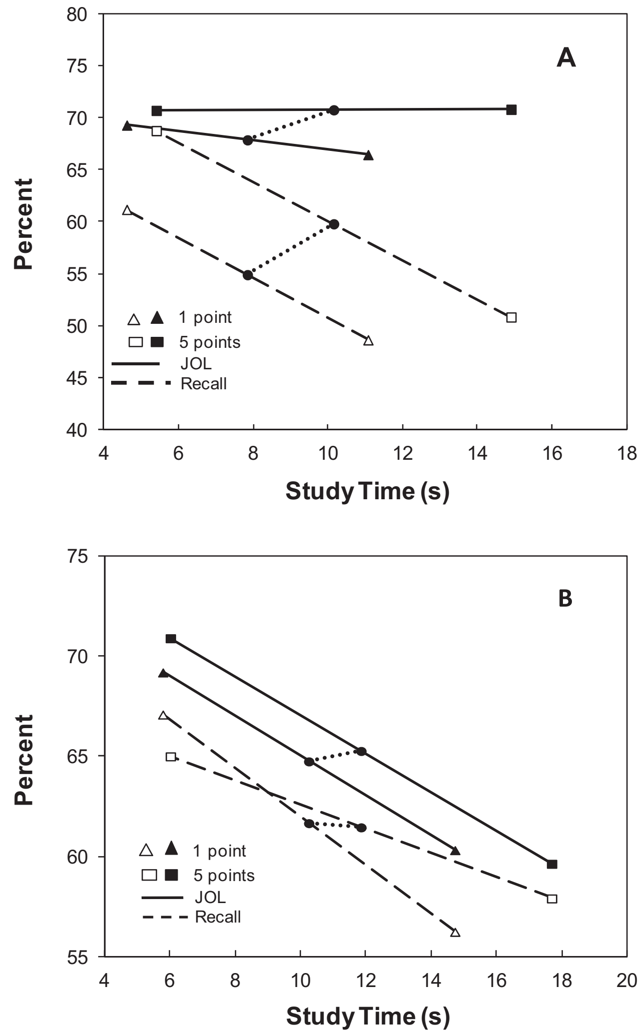

Figure 2. Mean judgment of learning (JOL) for fifth and sixth graders for below-median and above-median study time (in seconds) for each incentive level. Panel A: Related and unrelated items in separated blocks; each block contained low- and high-incentive items (Experiment 1). Panel B: Low- and high-incentive items in separated blocks; each block contained related and unrelated pairs (Experiment 2).

In comparing the means of the two incentive conditions, one can see that both JOLs and recall increased with ST. This result suggests sensitivity to goal-driven variation. However, whereas recall also yielded sensitivity to data-driven variation, decreasing with increasing ST for each incentive condition, JOLs evidenced little similar decrease. Let us examine ST allocation, JOLs, and recall for each type of regulation in turn.

Goal-driven regulation. Across both study blocks, children invested significantly more ST in 5-point items (10.2 s) than in 1-point items (7.8 s), $t(39) = 5.84$, $p < .0001$, Cohen’s $d = 0.92$. Of the 40 participants, 33 exhibited this trend, $p < .0001$, by a binomial test. In parallel, JOLs were also higher for 5-point items (70.8) than for 1-point items (67.9), $t(39) = 2.82$, $p < .01$, $d = 0.45$. This trend was observed for 29 out of the 40 participants, $p < .005$, by a binomial test. Thus, consistent with the MC model, the children regulated their ST in accordance with the goal of increasing their winnings, and JOLs increased with increasing ST. Recall was also somewhat better for 5-point items (59.8%) than for 1-point items (54.9%), $t(39) = 1.81$, $p < .08$, $d = 0.29$.

These results indicate that the children made a strategic use of ST, allocating more ST to the high-incentive items, and this regulation was effective to some extent in enhancing their recall. The increased ST was accompanied by higher JOLs so that JOLs made at the end of a study trial increased with increasing ST, consistent with the MC model.

Data-driven regulation. The effects of data-driven regulation were examined by comparing JOLs and recall between short and long STs (see Figure 2), as was done in previous studies (Koriat et al., 2006; 2009b). The results failed to yield evidence for sensitivity to data-driven variation. A two-way analysis of variance (ANOVA), Incentive (1 vs. 5 points) × ST (long vs. short) for JOLs yielded $F(1, 39) = 7.98$, mean square error (MSE) = 41.84, $p < .01$, $\eta_p^2 = .17$, for incentive, but $F < 1$ for ST, and no interaction, $F(1, 39) = 2.82$, $MSE = 30.85$, $p < .11$, $\eta_p^2 = .07$.

We also examined the within-participant Pearson correlation across items between ST and JOLs, using the full range of STs. This correlation averaged −.05 for 5-point items and −.12 for 1-point items, but only for the 1-point items was the correlation significant, $t(39) = 2.02$, $p < .05$, $d = 0.32$.

These results are particularly surprising in light of the fact that recall decreased with increasing ST for each of the incentive levels. Thus, a two-way ANOVA for recall yielded $F(1, 39) = 3.29$, $MSE = 291.04$, $p < .10$, $\eta_p^2 = .08$ for incentive; $F(1, 39) = 10.58$, $MSE = 876.34$, $p < .005$, $\eta_p^2 = .21$ for ST; and $F < 1$ for the interaction. Across both incentives, recall was higher for short-ST items (65.0%) than for long-ST items (49.7%). This trend was significant for the 5-point incentive condition, $t(39) = 3.48$, $p < .005$, $d = 0.55$, as well as for the 1-point incentive condition, $t(39) = 2.04$, $p < .05$, $d = 0.24$.[2]

With regard to JOL accuracy, the within-person JOL–recall gamma correlation (Nelson, 1984) averaged .29, $t(36) = 3.81$, $p < .001$, $d = 0.63$, for 5-point items and .42, $t(36) = 6.07$, $p < .0001$, $d = 0.99$, for 1-point items. It was .36, $t(39) = 7.41$, $p < .0001$, $d = 1.17$, across all items.

The JOL results of Experiment 1 are unexpected, first because previous studies yielded evidence for JOL's sensitivity to data-driven regulation even for third graders (Hoffmann-Biencourt et al., 2010; Koriat et al., 2009b) and for fourth graders, even when all items were unrelated (Koriat et al., 2009a). Second, because unlike JOLs, recall performance did decrease with ST for same-incentive items. In contrast, with regard to goal-driven regulation, the results confirmed that fifth and sixth graders allocate ST to items in accordance with the incentive awarded to the recall of these items and, in parallel, expect recall to be better for high-incentive items than for low-incentive items.

Two explanations may be offered for the failure of the children's JOLs to respond to data-driven regulation. The first is that perhaps the blocking of items by relatedness did not allow a sufficiently large variation between items within a block to permit children's JOLs to exhibit sensitivity to data-driven differences in ST. However, the average within-person variance in ST was about the same in this experiment as it was for fifth and sixth graders in Koriat et al.'s (2009b) study: In that study, there was no manipulation of incentive and the related and unrelated pairs were intermixed; the mean ST standard deviation across items was 4.6 s, whereas in this experiment it was 3.4 s for the block of related pairs and 5.9 s for the block of unrelated pairs.

A second explanation is that children find it difficult to respond to both goal-driven and data-driven variation within the same situation. Bringing differential incentives into focus resulted in enhanced sensitivity to goal-driven variation, but impeded sensitivity to data-driven variation. As a result, although ST was indeed diagnostic of recall, the children failed to make use of it as a cue for JOLs. It is interesting to note that this is the pattern that was observed for the younger children in Koriat et al.'s study (2009b): In that study, ST was a valid predictor of recall even for first and second graders, decreasing with increased ST, but children at these ages did not seem to make use of ST as a cue for JOLs. Thus, perhaps children's responsiveness to goal-driven regulation in the present study impaired their sensitivity to the feedback that they could derive from ease of learning.

Experiment 2 was designed to help distinguish between these two accounts by creating a condition that was expected to increase children's sensitivity to data-driven regulation.

Experiment 2

Experiment 2 was similar to Experiment 1 except that incentive was manipulated between blocks, whereas item relatedness varied within block. As in Experiment 1, children were given instructions that emphasized the importance of regulating learning according to the incentive associated with each item. However, the items were divided between two blocks according to incentive so that each block included a mixture of unrelated and related pairs, but all items in one block received a high incentive, whereas all those in the other block received a low incentive. This division was expected to increase sensitivity to data-driven variation relative to goal-driven variation in ST.

Method

Participants. Participants were 40 children drawn from the same population. The sample consisted of 20 fifth graders (mean age 10.6 years) and 20 sixth graders (mean age 11.5 years), with an overall mean age of 11.1 years.

Materials and procedure. The materials and procedure were the same as in Experiment 1, but the assignment of the items to the two blocks was such that each block included six related pairs and six unrelated pairs. All items in one block were associated with a 1-point incentive, whereas those in the other block were associated with a 5-point incentive, with the assignment counterbalanced across participants. Two items were used for practice in each list.

[2] Note that in calculating the within-person JOL–recall or ST–recall correlations, some participants had to be removed because they exhibited 100% recall in one of the incentive conditions. The effective number of participants in each case can be inferred from the reported degrees of freedom.

Results and Discussion

The results for JOLs and recall are plotted in Figure 2 (Panel B). It can be immediately seen that now the results clearly disclose the inverse relationship between ST and JOL that is characteristic of data-driven regulation. Recall reflects this pattern as well. However, the effects of goal-driven regulation disappeared almost entirely for both JOL and recall. We examine the results beginning first with those pertaining to data-driven regulation.

Data-driven regulation. JOLs for short-ST items and long-ST items averaged 70.0 and 60.0, respectively, $t(39) = 5.33$, $p < .0001$, $d = 0.84$. The ST–JOL Pearson correlation averaged $-.28$ for 5-point items, $t(39) = 5.54$, $p < .0001$, $d = 0.88$, and $-.18$ for 1-point items, $t(39) = 3.20$, $p < .005$, $d = 0.50$. Thus, JOLs decreased with increasing STs irrespective of incentive.

A similar pattern was observed for recall: Recall averaged 66.0% for short-ST items, and 57.1% for long-ST items, $t(39) = 2.67$, $p < .05$, $d = 0.42$. The JOL–recall gamma correlation averaged .44, $t(34) = 6.53$, $p < .0001$, $d = 1.10$, for 5-point items, and .55, $t(34) = 8.79$, $p < .0001$, $d = 1.49$, for 1-point items. It was .48, $t(38) = 9.41$, $p < .0001$, $d = 1.51$, across all items.

Goal-driven regulation. In contrast to the results for data-driven regulation, the children exhibited little sensitivity to goal-driven variation in ST. Across both study blocks, they invested only slightly more ST in 5-point items (11.9 s) than in 1-point items (10.3 s), $t(39) = 1.11$, $p = .28$, $d = 0.17$. There was also little difference between 5-point items and 1-point items either in JOLs or in recall. JOLs averaged 65.3 for 5-point items and 64.8 for 1-point items, $t(39) = 0.36$, $p = .72$, $d = 0.06$. The respective means for recall were 61.5 and 61.7, $t(39) = 0.08$, $p = .94$, $d = 0.01$. Thus, there was no sign for goal-driven regulation of ST or for its expected effects on metacognitive judgments.

A comparison of the results of Experiments 1 and 2 brings to the fore the pliability of metacognitive judgments. Whereas JOLs in Experiment 1 exhibited sensitivity to goal-driven regulation but not to data-driven regulation, Experiment 2 indicated data-driven regulation but not goal-driven regulation. We conducted several analyses to obtain some insight into the processes underlying the differential sensitivity of JOLs to goal-driven versus data-driven variation. Consider the results presented in Figure 3. In Panel A, JOLs are plotted as a function of mean ST for 1-point and 5-point incentives in Experiments 1 and 2, thus depicting the effects of goal-driven regulation. In Panel B, in contrast, they are plotted as a function of mean ST for short (below-median) and long (above-median) ST, thus depicting the effects of data-driven regulation. Two observations are noteworthy. First, the extent of ST variation due to goal-driven regulation (i.e., incentive; Panel A) is larger in Experiment 1 (amounting to 2.4 s) than in Experiment 2 (1.6 s). In contrast, the opposite is observed for data-driven regulation (Panel B): The extent of variation that is due to data-driven regulation is larger in Experiment 2 (10.3 s) than in Experiment 1 (8.0 s). Thus, the regulation of ST was more tuned to goal-driven variation in Experiment 1 and to data-driven variation in Experiment 2. This difference in self-regulation, in itself, may explain why JOLs vary more strongly (positively) with incentive in Experiment 1 and more strongly (negatively) with data-driven ST in Experiment 2.

A second observation, however, suggests a difference in monitoring over and above that which can be accounted for by differences in regulation. This is reflected in the slopes of the ST–JOL functions: The slope for goal-driven regulation was steeper in Experiment 1, whereas the slope for data-driven regulation was steeper in Experiment 2. This pattern suggests that conditions that make one dimension of variation more salient than the other not only make self-regulation differentially sensitive to variations in the two dimensions but also make monitoring differentially sensitive to the output of that regulation. It is as if in making JOLs, variation in self-regulation along the salient dimension is weighted more heavily than variation along the nonsalient dimension.

In conclusion, the results of Experiment 1 and Experiment 2 disclose an "either or" pattern: When one type of variation is more salient, children respond to it, exhibiting insensitivity to the other type of variation. This pattern of results may be due to the opposite implications that variations in ST have for metacognitive judgments in the case of goal-driven variation versus data-driven variation.

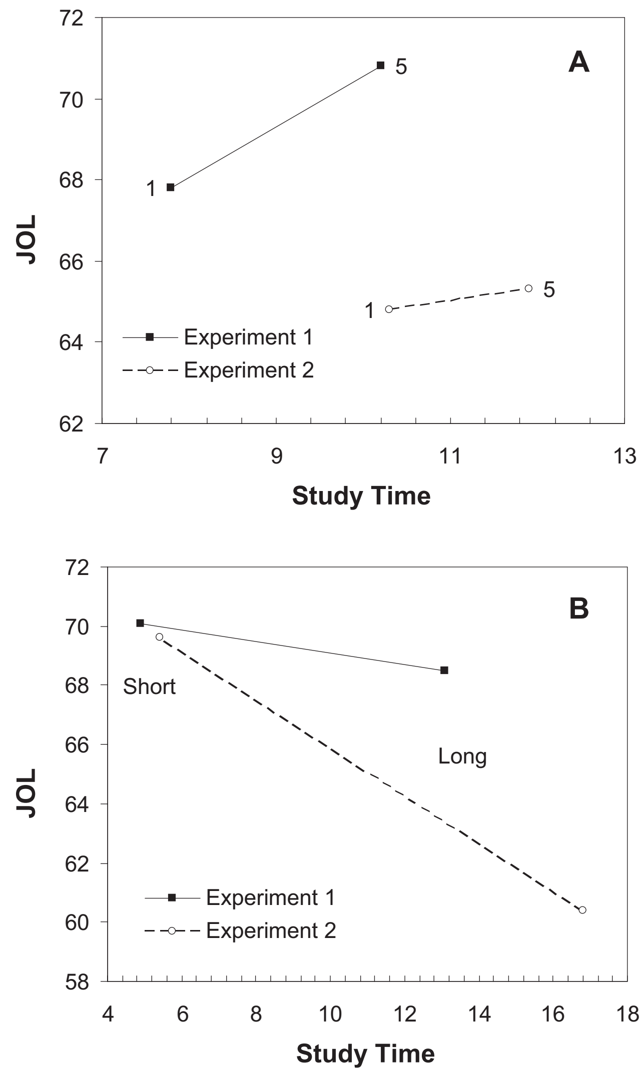

Figure 3. Mean judgment of learning (JOL) plotted separately for Experiment 1 and Experiment 2 as a function of the mean study time (in seconds) for two incentive levels (1 and 5; Panel A) and for below-median (short) and above-median (long) study times (Panel B).

Experiment 3

Experiment 3 was intended to provide a summary description of what seems to emerge from the results of Experiments 1 and 2. In Phase 1 of the experiment, we attempted to produce a more balanced emphasis on goal-driven and data-driven variation than in the previous two experiments. The design was similar to that used by Koriat et al. (2006, Experiment 5) with college students. We expected that the results would fail to yield metacognitive sensitivity to the effects of both types of variation. Therefore, Phase 2 of the experiment was intended to confirm that the same participants do respond to each dimension of variation when variation in the other dimension is held constant as found in the previous two experiments.

Method

Participants. Participants were 40 children drawn from the same population as in the previous experiments. The sample consisted of 20 fifth graders (mean age 10.7 years) and 20 sixth graders (mean age 11.6 years), with an overall mean age of 11.1 years.

Materials and procedure. The materials for Phase 1 of the experiment were the same as in Experiment 1. The items, however, were intermixed within a single list. Half of the items in each relatedness category were associated with a 1-point incentive, and the other half with a 5-point incentive.

For Phase 2, the children in each grade were divided randomly between the constant- and differential-incentive conditions with an equal number in each condition. For the constant-incentive condition, a new study list of 12 pairs was used (and two practice pairs) including six related and six unrelated pairs. Participants were told that the recall of each pair was worth 3 points. Participants in the differential-incentive condition were presented with a list consisting of 12 new pairs, all unrelated (and two practice pairs). Half of these pairs were associated with a 1-point incentive and half with a 5-point incentive, with the assignment counterbalanced across participants. The procedure of Phase 2 was otherwise the same as that of Phase 1.

Results and Discussion

Analyses of the results of Phase 1. The results from Phase 1 of the experiment are plotted in Figure 4. Participants allocated significantly more ST to the 5-point items (8.6 s) than to the 1-point items (7.4 s), $t(39) = 3.02$, $p < .005$, $d = 0.48$. However, JOLs did not differ significantly as a function of incentive, averaging 66.7 for 5-point items and 65.3 for 1-point items, $t(39) = 1.31$, $p = .20$, $d = 0.21$. Recall also failed to vary significantly with incentive: It averaged 57.1% for 5-point items and 54.5% for 1-point items, $t(39) = 1.01$, $p = .32$, $d = 0.16$. Thus, the children evidenced goal-driven regulation, allocating longer ST to the high-incentive than to the low-incentive items, but JOL failed to exhibit sensitivity to the feedback from that regulation.

In contrast, there was evidence indicating that JOLs were affected by data-driven variation, being higher for short-ST items (69.0) than for long-ST items (63.0), $t(39) = 3.05$, $p < .005$, $d = 0.48$. The ST–JOL Pearson correlation averaged –.11 for 5-point items, $t(39) = 2.17$, $p < .05$, $d = 0.34$, and $-.10$ for 1-point items, $t(39) = 2.00$, $p < .06$, $d = 0.32$. Thus, JOLs decreased with increasing STs, consistent with the pattern expected for data-driven regulation.

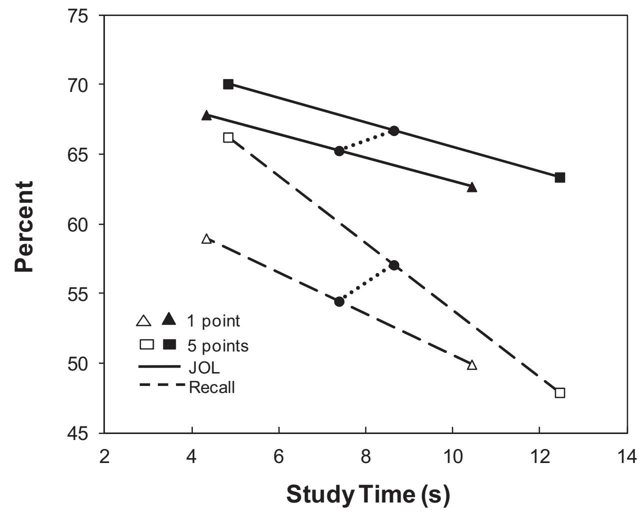

Figure 4. Mean judgment of learning (JOL, solid lines) and recall (broken lines) for fifth and sixth graders for below-median and above-median study time for each incentive level. Plotted also (dotted lines) are mean JOL and recall as a function of mean study time (in seconds) for each incentive level (Experiment 3, Phase 1).

Recall was also better (62.6%) for short-ST items than for long-ST items (48.9%), $t(39) = 5.28$, $p < .0001$, $d = 0.83$. The ST–recall gamma correlation averaged –.28 for 5-point items, $t(38) = 5.16$, $p < .0001$, $d = 0.82$, and –.19 for 1-point items, $t(38) = 3.53$, $p < .005$, $d = 0.57$. The JOL–recall gamma correlation averaged .64, $t(38) = 13.57$, $p < .0001$, $d = 2.18$ for 5-point items and .61 for 1-point items, $t(38) = 13.52$, $p < .0001$ $d = 2.16$. Across all items the JOL–recall gamma was .62, $t(39) = 18.61$, $p < .0001$, $d = 2.94$.

In sum, the results of Phase 1 yielded evidence for metacognitive judgments being based on data-driven regulation of study effort. However, although relatively more ST was allocated to the high-incentive items, there was little indication for the expected effects of goal-driven regulation of ST on JOLs and recall.

Analyses of the results of Phase 2. In Phase 2 of the experiment, the constant-incentive participants exhibited sensitivity to data-driven variation, as was found in Experiment 2 (see Figure 5, Panel A): JOLs were higher for below-median ST (76.6) than for above-median STs (62.1), $t(19) = 4.40$, $p < .001$, $d = 0.98$. Recall was also better for below-median STs (66.3%) than for above-median STs (54.9%), although the difference was not significant, $t(19) = 1.36$, $p = .19$, $d = 0.30$. The ST–JOL correlation averaged –.26, $t(19) = 5.24$, $p < .0001$, $d = 1.17$. Thus, JOLs decreased with increasing STs, consistent with the pattern expected for data-driven regulation. The JOL–recall gamma correlation averaged .42, $t(18) = 5.33$, $p < .0001$, $d = 1.22$.

The results for the differential-incentive condition (see Figure 5, Panel B) yielded little evidence of data-driven regulation. Across incentives, JOLs averaged 51.6 for below-median STs and 51.0 for above-median STs, $t(19) = 0.24$, $p = .82$, $d = 52.92$. The results for recall were surprising: Across incentives, recall actually tended

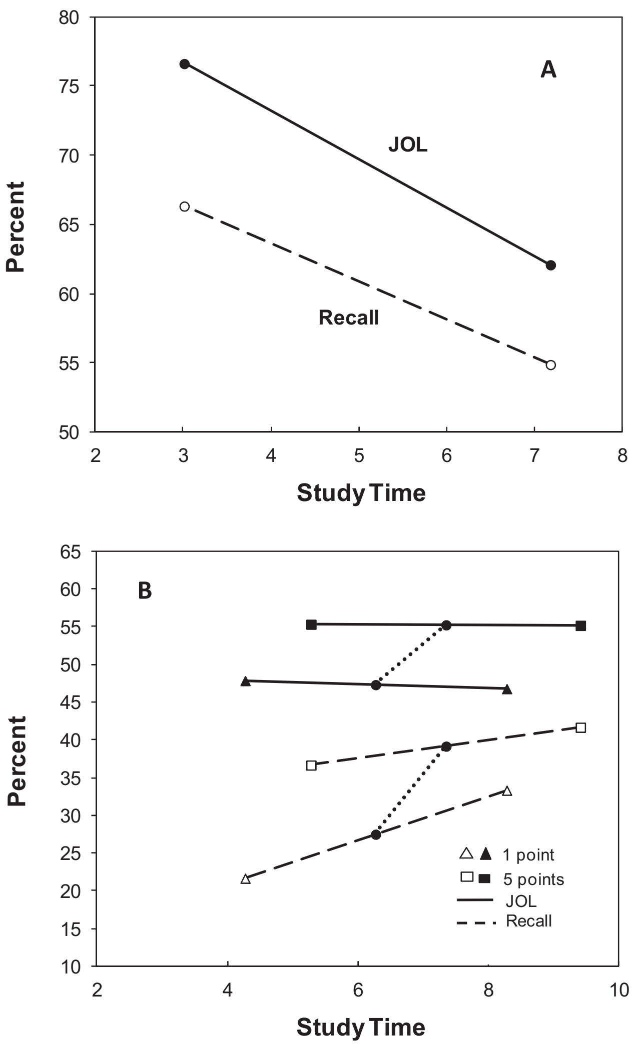

Figure 5. Mean judgment of learning (JOL, solid lines) and recall (broken lines) for fifth and sixth graders for below-median and above-median study time (in seconds) for the constant-incentive group (Panel A) and for each level of the differential-incentive group (Panel B). Plotted also (dotted lines) in Panel B are the means of JOL and recall as a function of mean study time for each incentive level (Experiment 3, Phase 2).

to increase from short STs (29.17) to long STs (37.50), $t(19) = 1.88$, $p < .08$, $d = 0.42$.

As for goal-driven regulation, participants invested more ST in 5-point items (7.4 s) than in 1-point items (6.3 s), but the difference was not significant, $t(19) = 1.15$, $p = .26$, $d = 0.26$. JOLs, in contrast, were sensitive to incentive averaging 55.2 for 5-point items and 47.3 for 1-point items, $t(19) = 3.02$, $p < .01$, $d = 0.68$. Recall was lower overall than in the previous experiments, because all the items were unrelated, but nevertheless it also yielded an effect of incentive, averaging 39.2% for 5-point items and 27.5% for 1-point items, $t(19) = 2.33$, $p < .05$, $d = 0.52$.

In sum, Experiment 3 (Phase 1) was similar in design to Experiment 5 of Koriat et al. (2006), but yielded evidence only for the effects of data-driven variation. The results suggest that the metacognitive monitoring of fifth and sixth graders is not mature enough to incorporate the effects of goal-driven variation: Although the children exerted goal-driven regulation of ST, their metacognitive judgments did not exhibit sensitivity to the ensuing variation in ST. The results from Phase 2 clearly indicated that this was not because these children could not respond to goal-driven variation, but possibly because they could not respond *simultaneously* to both types of variation.

Experiment 4

The results presented so far indicate that fifth and sixth graders can exhibit sensitivity to both goal-driven and data-driven variation in ST but cannot respond simultaneously to both types of variation in the same task. College students, in contrast, were able to react differentially to the two types of variation, demonstrating a positive ST–JOL relationship for goal-driven variation and a negative relationship for data-driven variation. In Experiment 4, we included a group of ninth graders in order to trace the developmental trajectory of this ability.

Method

Participants. Participants were 20 ninth graders (mean age 14.5 years) drawn from one of the schools used in the previous experiments that has a junior high school within the same school complex.

Materials and procedure. The materials and procedure were the same as in Experiment 3 (Phase 1).

Results and Discussion

We begin by reporting the results for the ninth graders. We then compare the three age groups—fifth and sixth graders, ninth graders, and college students (based on Koriat et al., 2006) in terms of several variables that can shed light on the underlying developmental changes.

Analysis of the results for 9th-graders. The results for ninth graders (Figure 6) were more similar to those of college students (Figure 1) than to those of the primary school children (Experiment 3, Phase 1; Figure 4), evidencing sensitivity to both goal-driven and data-driven regulation of ST within the same task. With regard to goal-driven regulation, participants allocated significantly more ST to the 5-point items (13.1 s) than to the 1-point items (10.2 s), $t(19) = 4.00$, $p < .001$, $d = 0.89$. In parallel, JOLs were higher for the 5-point items (64.4) than for the 1-point items (59.2), $t(19) = 4.23$, $p < .0005$, $d = 0.94$, and recall evidenced the same trend (the respective means were 67.1% and 62.5%), $t(19) = 1.81$, $p < .09$, $d = 0.41$. These results are consistent with the MC model.

At the same time, the results yielded evidence for sensitivity to data-driven variation, consistent with the CM model: JOLs were higher for short-ST items (67.9) than for long-ST items (55.7), $t(19) = 4.23$, $p < .0005$, $d = 0.94$. The ST–JOL correlation averaged –.39 for 5-point items, $t(19) = 6.84$, $p < .0001$, $d = 1.53$, and –.30 for 1-point items, $t(19) = 4.94$, $p < .0001$, $d = 1.11$. Recall was also higher for short-ST items (70.0%) than for long-ST items (59.6%), $t(19) = 2.25$, $p < .05$, $d = 0.50$. The JOL–recall gamma correlation averaged .54, $t(18) = 6.71$, $p < .0001$, $d = 1.50$, for 5-point items and .61 for 1-point items, $t(18) = 7.25$, $p < .0001$, $d = 1.66$. Across all items, the JOL–recall gamma correlation averaged .54, $t(19) = 7.34$, $p < .0001$, $d = 1.64$.

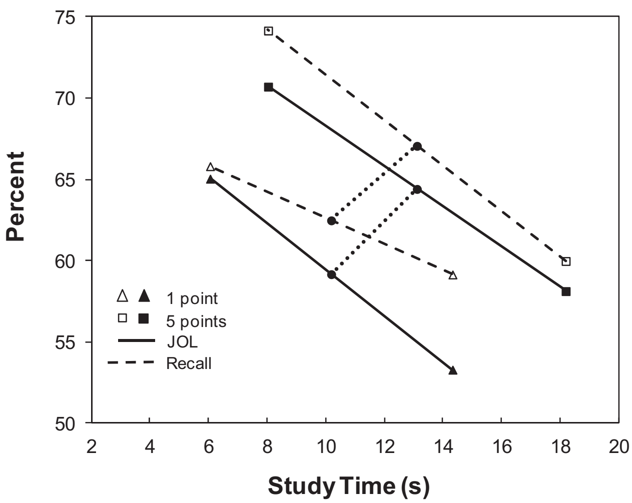

Figure 6. Mean judgment of learning (JOL, solid lines) and recall (broken lines) for ninth graders for below-median and above-median study time (in seconds) for each incentive level. Plotted also (dotted lines) are mean JOL and recall as a function of mean study time for each incentive level (Experiment 4).

In sum, the results for ninth graders disclose an adult-like pattern, evidencing simultaneous sensitivity to data-driven and goal-driven variation. The similarity between the pattern of results for JOLs and recall in Figure 6 is impressive, suggesting that the effects of STs on JOLs capture faithfully the respective effects on recall. These results strongly suggest that the metacognitive ability to respond differentially and simultaneously to the implications of data-driven and goal-driven regulation develops between the sixth and ninth grades. We shall now examine more closely the developmental changes that take place across the three age groups: fifth and sixth graders, ninth graders, and college students.

A comparison across the three age groups. Table 1 summarizes some of the developmental trends suggested by the results. The first two columns present the mean within-person ST–JOL correlation for the low-incentive and high-incentive items for the three age groups. The data were taken from Experiment 3 (Phase 1) for fifth and sixth graders, from Experiment 4 for ninth graders, and from Koriat et al. (2006, Experiment 5, 1st presentation) for college students. The correlations exhibit a monotonic increase with age for both low-incentive and high-incentive items, $F(2, 73) = 9.41$, $MSE = 0.08$, $p < .001$, $\eta_p^2 = .20$, and $F(2, 73) = 16.27$, $MSE = 0.08$, $p < .0001$, $\eta_p^2 = .31$, respectively, suggesting an age-related increase in the reliance of JOLs on data-driven variation in ST. As shown in Table 1, post hoc analyses of the difference among the groups by Scheffe's test revealed significant differences between the younger group and the other two groups, both $p < .05$, and no difference between the ninth graders and the young adults. However, hierarchical linear models (HLM; Proc Glimmix macro of SAS Version 9.2) indicated a significant difference between the two older groups in the slopes of the functions relating JOLs to ST. When these slopes were calculated for each participant, significant interactive effects were found, indicating that the slopes were steeper for college students than for the ninth graders for the low-incentive level, $t(397) = 5.3$, $p < .0001$, as well as for the high-incentive level, $t(543) = 6.89$, $p < .0001$. These results supplement the finding of Koriat et al. (2009b) and Hoffmann-Biencourt et al. (2010). They found a stronger decrease in JOLs with ST for third through sixth graders than for first and second graders. The present results, however, suggest increased sensitivity of JOLs to the feedback from data-driven regulation even beyond the primary school years.

Table 1 also presents the gamma correlations between incentive and JOL for related and unrelated pairs. For the related pairs, the correlations did not differ from zero for any of the age groups. For the unrelated pairs, in contrast, the correlations were significant for all age groups and tended to increase with age, but the differences among the groups were not significant, either by Scheffe's post hoc test or by HLM.

The critical property that distinguishes college students from primary school children, however, lies in the ability to respond simultaneously and differentially to data-driven variation and goal-driven variation rather than in the ability to respond to each of them alone. The following analyses were carried out in order to bring to the fore the qualitative change that takes place with age in the combined effects of the two types of variation. In the first analysis, for each participant, the effect size of goal-driven regulation on JOLs was calculated by subtracting mean JOLs for low-incentive items from mean JOLs for high-incentive items and dividing the difference by the participant's standard deviation of JOLs. In parallel, the effect size of data-driven regulation on JOLs was calculated by subtracting mean JOLs for below-median ST items from mean JOLs for above-median ST items and dividing the difference by the participant's standard deviation of JOLs. The ability of learners to respond to both types of variation is best reflected in the sum total of the two effect sizes for each participant. It can be seen in the last column of Table 1 that the sum total

Table 1
Mean Correlations and Effect Sizes for the Three Age Groups

	Study time–judgment of learning		Incentive–judgment of learning		Effect size		
Age group	Low incentive	High incentive	Related pairs	Unrelated pairs	Goal-driven	Data-driven	Goal-driven + data-driven
Fifth and sixth graders	-0.10_{a}^{*}	-0.11_{a}^{*}	0.03_{a}	0.12_{a}^{*}	0.13_{a}	0.56_{a}^{*}	0.68_{a}^{*}
Ninth graders	-0.30_{b}^{***}	-0.39_{b}^{***}	0.13_{a}	0.24_{a}^{*}	0.62_{b}^{***}	$0.89_{a,b}^{**}$	1.51_{b}^{***}
Undergraduates	-0.45_{b}^{***}	-0.56_{b}^{***}	0.12_{a}	0.31_{a}^{***}	$0.40_{a,b}^{**}$	1.49_{b}^{***}	1.90_{b}^{***}

Note. The subscripts a and b indicate significant group differences by Scheffe's post-hoc test.
$^{*}p \leq .05$; $^{**}p \leq .001$; and $^{***}p < .0001$ for the difference from zero.

increases monotonically with age. One-way ANOVAs conducted for each of the three effect size measures yielded, $F(2, 73) = 5.56$, $MSE = 0.92$, $p < .01$, $\eta_p^2 = .13$ for data-driven regulation; $F(2, 73) = 3.36$, $MSE = 0.51$, $p < .05$, $\eta_p^2 = .08$ for goal-driven regulation; and, $F(2, 73) = 9.31$, $MSE = 1.09$, $p < .0005$, $\eta_p^2 = .20$ for the effect size of both data-driven and goal-driven regulation. The results of the Scheffe's post hoc (see Table 1) indicate significant differences between the younger group and either one or both of the two older groups.

The developmental trend in the ability to respond to the two dimensions of variability at the same time was also confirmed by an analysis of the estimated proportion of variance in JOLs that is accounted for by goal-driven and data-driven effects. The estimates are based on two-way ANOVAs on JOLs, Incentive (low vs. high) × ST (below median vs. above median), for each age group. The percentage of variance accounted for by both effects together was 4.6% for the fifth and sixth graders, 16.3% for the ninth graders, and 48.5% for the college students.

In conclusion, the results of Experiment 4 suggest that unlike fifth and sixth graders, ninth graders are not only capable of exercising data-driven and goal-driven regulation in the same task but also possess the metacognitive sophistication needed to take into account the opposite effects of the two types of regulation in monitoring their own learning.

Experiment 5

Experiment 5 was designed to gain some insight into the metacognitive deficiency underlying the failure of fifth and sixth graders' JOLs to respond to the effects of both data-driven and goal-driven regulation for the same item. In the previous experiments, fifth and sixth graders exhibited sensitivity to each of these two types of regulation in making recall predictions, so what prevented them from doing so when both types of regulation occurred within the same task? We proposed that the difficulty derived from the fact that ST represents the *joint* output of both regulations. Thus, a child (or an adult for that matter) is faced with a task similar to that underlying the logic of analysis of variance: to *partition* the observed total variance into components attributable to two different sources of variation. We hypothesized that this partitioning is necessary for drawing the opposite implications that the two types of variation have for JOLs. This proposal implies that the metacognitive deficiency of primary school children derives specifically from a process that intervenes between (a) the regulation of ST according to data-driven and goal-driven variation and (b) the computation of metacognitive judgments that takes into account each of the two sources. Indeed, the results of Experiment 3 (Phase 1) indicated that fifth and sixth graders were able to regulate their ST according to *both* data-driven and goal-driven regulation in the same task. The results of Phase 2 indicated further that their JOLs were sensitive differentially to the two types of variation when these variations occurred separately. Therefore, the failure of these children to take into account both types of variation must lie in the differential attribution of variations in ST to their separate sources (see Koriat & Nussinson, 2009).

To illustrate the implications of this analysis, we consider two items: a difficult-to-encode item that is associated with a low incentive and an easy-to-encode pair that is associated with a high incentive. The results suggest that school children (Experiments 3 and 4) are successful in regulating ST in keeping with *both* goal-driven and data-driven regulation, perhaps investing the same amount of ST in both items. When it comes to monitoring, however, primary school children seem to encounter a difficulty in responding to the joint effects of *their* regulation. Although the same amount of ST might be allocated to the two items, learners ought to assign higher JOLs to the high-incentive related pair than to the low-incentive unrelated pair. Indeed, inspection of Figure 1 (college students) and of Figure 6 (ninth graders) indicates that JOLs differed markedly for items with the same ST depending on the specific combination of data-driven and goal-driven sources of ST. The discrepancy exhibited by children between the results for regulation and those for monitoring suggests that the ST invested in an item does not carry (or retain) a stamp of its source: Children do not keep a record of what proportion of their ST derives from their data-driven regulation and what proportion is due to their goal-driven regulation.

In Experiment 5, we examined this idea by exploring the effects of a training procedure designed to help children partition differences in ST according to their source. A before–after design was used, with the partitioning procedure (Phase 2) intervening between two phases. Phase 1 (pretraining) involved a similar procedure to that used in Phase 1 of Experiment 3 and in Experiment 4, and the procedure in Phase 3 (posttraining) was also the same, except that a new study list was used.

In Phase 2 (training) of the experiment, children were first allowed to invest as much ST as they wished in each item. Only then was the incentive associated with the item announced, and the children had the option of investing more ST in the item or ending study and moving to the next item. We assume that the first part of ST reflects data-driven regulation, whereas the second part reflects goal-driven regulation. We examined first, whether under these conditions, children's JOLs would be sensitive to both data-driven and goal-driven regulation. Second, we examined whether the effects of this procedure would transfer to the situation in which both types of regulation occur at the same time.

Method

Participants. The sample consisted of 30 fifth graders (mean age 10.8 years) and 30 sixth graders (mean age 11.6 years), with an overall mean age of 11.2 years. They were drawn from the same population as in the previous experiments with this age group.

Materials. The study items were 46 word pairs. They included the 24 word pairs that had been used in the previous experiments, and new items taken from the same pool of items and chosen to include 23 related pairs and 23 unrelated pairs. Six additional pairs were used for practice, two word pairs for each phase. The pairs were divided into three lists, one for the pretraining phase (15 pairs), one for the training phase (15 pairs), and one (16 pairs) for the posttraining phase.

Procedure. The children received instructions about incentives as had been done in the previous experiments. The procedure of the pretraining phase was identical to that of Phase 1 of Experiment 3. In the second (training) phase, each study pair was first presented without its associated incentive value. The children were instructed to study the pair as long as they needed. They indicated end of study by pressing a box labeled "How many stars?" The word pair and the box then disappeared, and one star or five blinking stars appeared with two boxes below marked "Study more" and "Continue." Pressing the "study-more" box displayed the pair again, and the star indication remained on the screen. The child could then continue studying the item as long as needed, clicking the "continue" box when done. Upon pressing the "continue" box, the word pair was replaced by the JOL scale.

The procedure for the posttraining phase was the same as that of the pretraining phase but with a new list of items. Following this phase and after a filler task, a test phase took place, including all cue words from the three study phases, randomly ordered.

Results and Discussion

We shall first present the results for Phase 1 in order to examine later the effects of the partitioning procedure.

Phase 1 (pretraining). Figure 7 (Panel A) presents the results for this phase in the same format as in the previous experiments. With regard to goal-driven regulation, children invested significantly more ST in 5-point items (10.8 s) than in 1-point items (8.6 s), $t(59) = 5.29$, $p < .0001$, $d = 0.68$. However, the respective means for JOLs were 65.4 and 63.6 and did not differ significantly, $t(59) = 1.55$, $p = .13$, $d = 0.20$. Recall also failed to yield a significant effect of incentive: It averaged 56.5% for 5-point items and 54.2% for 1-point items, $t(59) = 0.75$, $p = .44$, $d = 0.10$.

In contrast, JOLs were responsive to data-driven regulation: They were higher for below-median STs (66.7) than for above-median STs (62.4), $t(59) = 2.93$, $p < .005$, $d = 0.38$. The ST–JOL Pearson correlation averaged $-.16$ for 5-point items, $t(59) = 3.11$, $p < .005$, $d = 0.40$, and $-.12$ for 1-point items, $t(59) = 2.25$, $p < .05$, $d = 0.29$. Recall for short-ST items (56.6%) did not differ from recall for long-ST items (54.0%), $t(59) = 0.87$, $p = .39$, $d = 0.11$. The JOL–recall gamma correlation averaged .34, $t(54) = 5.27$, $p < .0001$, $d = 0.71$, for 5-point items and .36 for 1-point items, $t(54) = 5.47$, $p < .0001$, $d = 0.74$.

In sum, the results for Phase 1 were similar to those of Experiment 3 (Phase 1): The children regulated the allocation of ST in accordance with the differential incentives, but their JOLs did not increase significantly with incentive. In contrast, JOLs were sensitive to data-driven regulation, decreasing with increasing STs.

Phase 2 (training). Figure 8 presents the results for the partitioned ST allocation for related and unrelated pairs. In this figure, ST is partitioned into two components. The first component, prior to the announcement of incentive, is assumed to reflect data-driven regulation. The second component, the added ST in response to the announced incentive, is assumed to reflect goal-driven regulation. Participants spent initially 6.9 s on average studying each item, 5.9 s on the related pairs and 8.0 s on the unrelated pairs, $t(59) = 4.64$, $p < .0001$, $d = 0.60$. Additional ST was then allocated to the 5-point items (4.6 s) than to the 1-point items (1.0 s), $t(59) = 6.51$, $p < .0001$, $d = 0.84$. Thus, the total amount of time invested in the study of each item averaged 11.3 s for 5-point items and 8.1 s for 1-point items, $t(59) = 5.25$, $p < .0001$, $d = 0.70$.

Note that the effects of incentive were stronger for the unrelated pairs than for the related pairs. Thus, an Incentive × Relatedness (related vs. unrelated) ANOVA on the added (goal-driven) ST yielded $F(1, 59) = 54.74$, $MSE = 17.15$, $p < .0001$, $\eta_p^2 = .48$, for

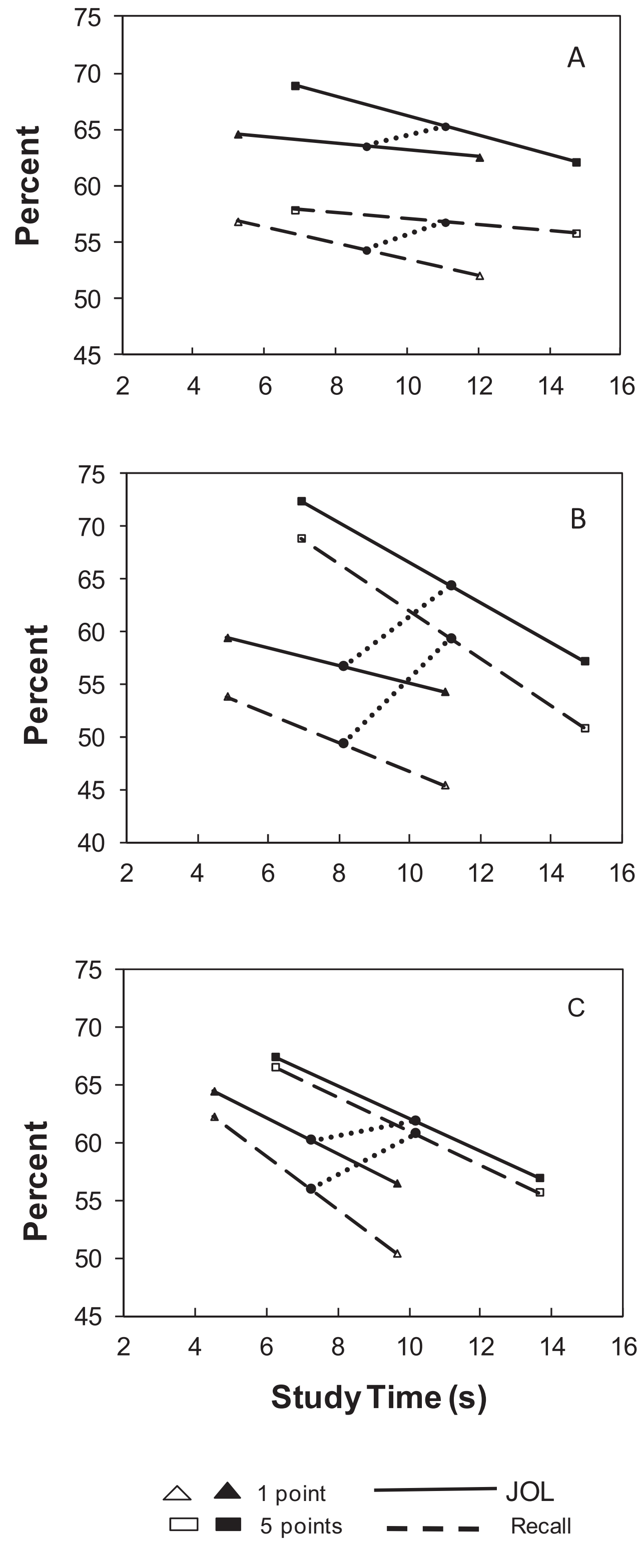

Figure 7. Mean judgment of learning (JOL, solid lines) and recall (dotted lines) for fifth and sixth graders for below-median and above-median study time (in seconds) for each incentive level (Experiment 5). Panel A shows pretraining phase; Panel B shows training phase; and Panel C shows posttraining phase.

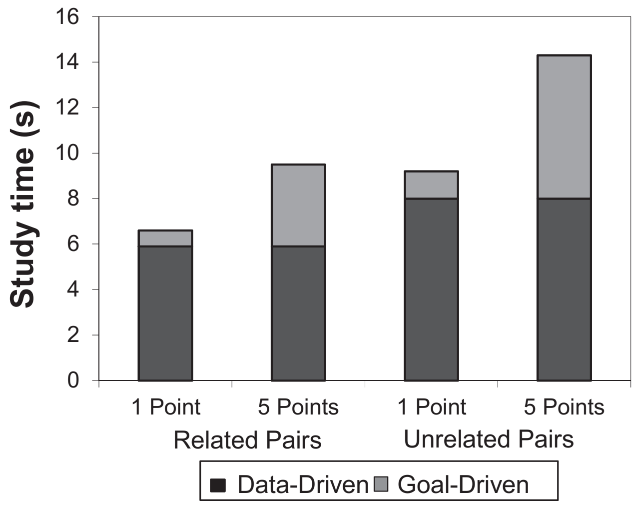

Figure 8. The study time allocation (in seconds) in Phase 2 of Experiment 5 partitioned according to initial (data-driven) and added (goal-driven) effects in Experiment 5 as a function of item relatedness (related vs. unrelated pairs) and incentive (1 vs. 5 points).

incentive; $F(1, 59) = 17.18$, $MSE = 8.89$, $p < .0001$, $\eta_p^2 = .23$, for relatedness; and $F(1, 59) = 17.40$, $MSE = 4.24$, $p < .0001$, $\eta_p^2 = .23$, for the interaction. A similar interaction was reported also by Koriat et al. (2006). It was interpreted in terms of Kahneman's (1973) proposal that the effort invested in a task is constrained by task difficulty: People cannot try as hard in a relatively easy task as they do when the task becomes more demanding.[3]

We turn next to examination of JOLs and recall in Phase 2 (Figure 7, Panel B). JOLs were higher (64.7) for 5-point items than for 1-point items (56.6), $t(59) = 4.76$, $p < .0001$, $d = 0.61$. Recall also increased with increasing incentive: It was higher for 5-point items (60.0%) than for 1-point items (49.0%), $t(59) = 3.83$, $p < .0005$, $d = 0.49$. JOLs were also sensitive to data-driven regulation: They were higher for short-ST items (65.6) than for long-ST items (55.8), $t(59) = 6.17$, $p < .0001$, $d = 0.80$. The ST–JOL correlation averaged –.39 for 5-point items, $t(59) = 8.11$, $p < .0001$ $d = 1.05$, and –.16 for 1-point items, $t(59) = 3.64$, $p < .001$, $d = 0.47$. Recall was also higher for short-ST items (60.9%) than for long-ST items (48.1%), $t(59) = 3.62$, $p < .001$, $d = 0.47$. The JOL–recall gamma correlation averaged .40, $t(59) = 5.30$, $p < .0001$, $d = 0.74$, for 5-point items, and .44 for 1-point items, $t(59) = 5.77$, $p < .0001$, $d = 0.81$.

A comparison of the results of Phase 2 with those of Phase 1 confirmed that the effects of incentive as well as the effects of ST were significantly stronger for Phase 2 than for Phase 1. A three-way ANOVA, Phase (1 vs. 2) × Incentive (1 vs. 5 points) × ST (long vs. short) for JOLs yielded, $F(1, 59) = 10.62$, $MSE = 112.13$, $p < .005$, $\eta_p^2 = .15$, for the Phase × Incentive interaction, and $F(1, 59) = 6.82$, $MSE = 130.21$, $p < .05$, $\eta_p^2 = .10$, for the Phase × ST interaction. These interactions suggest that the partitioning procedure increased the effects of both goal-driven and data-driven variations on JOLs relative to Phase 1. A similar three-way ANOVA on recall yielded a near-significant Phase × Incentive interaction, $F(1, 59) = 3.94$, $MSE = 573.86$, $p < .06$, $\eta_p^2 = .06$, and a significant Phase × ST interaction, $F(1, 59) = 4.56$, $MSE = 667.16$, $p < .05$, $\eta_p^2 = .07$.

In sum, the interactions with Phase that were obtained for both JOL and recall suggest that in comparison to Phase 1, the partitioning procedure used in Phase 2 was successful in producing sensitivity to goal-driven variation, and strengthening sensitivity to data-driven variation.

Phase 3 (posttraining). We turn finally to the results of Phase 3 (Figure 7, Panel C), examining whether the effects of the partitioning procedure in Phase 2 transferred to the typical regulation of ST when the incentive associated with each item is indicated at the beginning of a study trial, as was done in all the previous experiments as well as in Phase 1 of the present experiment.

We consider first goal-driven regulation. The children invested more ST in 5-point items (9.8 s) than in 1-point items (7.1 s), $t(59) = 4.91$, $p < .0001$, $d = 0.63$, similar to what was found in Phase 1. The respective means for JOLs, however, were 62.1 and 60.3, $t(59) = 1.35$, $p = .18$, $d = 0.17$. Recall was 60.9% for 5-point items and 56.2% for 1-point items, $t(59) = 1.45$, $p = .15$, $d = 0.19$.

With regard to data-driven regulation, JOLs were higher for short ST items (65.6) than for long-ST items (56.8), $t(59) = 5.07$, $p < .0001$, $d = 0.65$. The JOL–ST correlation averaged –.30 for 5-point items, $t(59) = 6.58$, $p < .0001$, $d = 0.85$, and –.21 for 1-point items, $t(59) = 4.72$, $p < .0001$, $d = 0.61$. Recall was also higher for short-ST items (63.9%) than for long-ST items (53.2%), $t(59) = 3.72$, $p < .0005$, $d = 0.48$. The JOL–recall gamma correlation averaged .54, $t(49) = 7.91$, $p < .0001$, $d = 1.12$, for 5-point items and .51 for 1-point items, $t(49) = 7.79$, $p < .0001$, $d = 1.10$.

A three-way ANOVA on JOLs that included phase as a factor (1 vs. 3) yielded $F < 1$ for the Phase × Incentive interaction, indicating that the effects of incentive that were brought to the fore by the partitioning procedure in Phase 2 did not transfer to Phase 3. In contrast, the Phase × ST interaction was significant, $F(1, 59) = 4.47$, $MSE = 136.96$, $p < .05$, $\eta_p^2 = .07$, suggesting that the effects of the partitioning procedure did subsist as far as data-driven regulation is concerned.

Note that in previous studies, the ST–JOL correlation was found to increase with repeated presentations of the same list (Koriat, 1997; Koriat & Bjork, 2006a, 2006b; Koriat et al., 2006). This increase, however, was not obtained when different lists were used across study–test blocks (Koriat, 1997; Koriat & Bjork, 2006b). So perhaps the improvement in sensitivity from Phase 1 to Phase 3 should be taken seriously.

One final observation concerns the discrepancy between JOL and recall. In all previous experiments as well as in Phase 1 of the present study, JOLs were generally higher than recall among fifth and sixth graders. It is interesting that in Phase 2 as well as in Phase 3, JOLs are at about the same level as recall, as was found to be the case for ninth graders. It is difficult to know, however,

[3] A similar ANOVAs for the previous experiments indicated a significant or near-significant Incentive × Relatedness interaction in Experiment 1, $F(1, 39) = 10.53$, $MSE = 7.85$, $p < .005$, $\eta_p^2 = .21$; in Experiment 3 (Phase 1), $F(1, 39) = 3.37$, $MSE = 4.47$, $p < .08$, $\eta_p^2 = .08$; and in Experiment 4, $F(1, 19) = 3.67$, $MSE = 4.32$, $p < .08$, $\eta_p^2 = .17$. The interaction was not significant in Experiment 2, $F(1, 39) = 1.12$, $MSE = 14.43$, $p = .30$, $\eta_p^2 = .03$, but it was in the same direction as in the previous experiments.

whether the change in calibration observed between Phase 1 and Phases 2 and 3 is also a result of the partitioning procedure used in Phase 2.

In conclusion, the results of Experiment 5 are consistent with the partitioning hypothesis of the metacognitive deficiency that prevents fifth and sixth graders from responding to both data-driven and goal-driven variation within the same task. According to this hypothesis, although the children are relatively successful in regulating ST according to both data-driven and goal-driven effects, they fail to separate the two sources when it comes to metacognitive judgments. Indeed, the partitioning procedure used in the training phase brought to the fore the opposite effects of the two types of variation. However, the training procedure had only a limited transfer to the task used in Phase 3.

General Discussion

The MC and CM models discussed by Koriat et al. (2006; see Koriat, 2006) have conflicting implications for metacognitive judgments. On the one hand, there is evidence that metacognitive monitoring drives metacognitive control (Koriat & Goldsmith, 1996; Metcalfe & Finn, 2008; Thiede et al., 2003), consistent with the MC model. Also, studies of learning and memory have repeatedly indicated that memory performance increases with the amount of experimentally allocated ST (e.g., Zacks, 1969). Not surprisingly, learners, even 4-year-old children (O'Sullivan, 1993), hold the belief that increased effort yields better recall. At the same time, there is evidence that under self-regulated learning, more study time and more trials to acquisition are associated with poorer memory performance (Koriat, 2008; Koriat et al., 2006, 2009a, 2009b). The conceptual framework proposed by Koriat et al. (2006) attempted to reconcile between these two observations in terms of the distinction between goal-driven and data-driven regulation in self-paced learning. It was argued that both types of regulation may occur within the same task, which should present a challenge to learners' monitoring of their degree of mastery of each item. Adult learners exhibited sensitivity to both data-driven and goal-driven variations within the same situation, providing JOLs that mirrored the opposite effects of the two types of variation on recall.

The distinction between data-driven and goal-driven regulation bears some similarity to the distinction between reactive control and effortful control (see Derryberry & Rothbart, 1997; Eisenberg & Morris, 2002), which has figured prominently in developmental studies of emotional self-regulation (Baumeister & Vohs, 2004). Reactive control involves responding to external influences such as the narrowing of attention that occurs when facing a threatening situation. Effortful (or proactive) control, in contrast, involves the ability to initiate or inhibit action voluntarily in a planful way. Similarly, data-driven regulation involves an adjustment of study effort to the encoding requirements of an item in a bottom–up fashion. Goal-driven regulation, in contrast, is more effortful and deliberate, possibly requiring a greater degree of top-down executive control.

The assumption that in self-paced learning ST is typically data-driven implies a CM model in which JOLs are based on the feedback from memorizing effort. This assumption predicts a negative JOL–ST correlation, but we should stress that that correlation is also consistent with theories in which JOLs are assumed to drive ST (MC model). One version of such theories is that learners judge the difficulty of each item in advance of learning and regulate ST so as to compensate for differences in a priori item difficulty (Nelson & Leonesio, 1988). A more dynamic version is postulated by the discrepancy-reduction model (Dunlosky & Hertzog, 1998): Learners continue to monitor the increase in encoding strength of an item as more study time is invested and cease when a preset level has been reached. Normatively difficult items are assumed to require more study time to reach the same preset level than easier items.

We agree that both ST regulation and JOLs may be mediated by reliance on naïve theories about the effects of various item properties on recall (see Mueller et al., in press; Koriat, 1997). However, we argue that the allocation of ST to a given item in self-paced learning is typically data-driven, determined on line by the ease with which each learner manages to encode that particular item. Although items tend to differ reliably across learners in the ease with which they are committed to memory, JOLs mirror the specific idiosyncratic experience that a learner gains from studying an item. Indeed, in Koriat (1997), interparticipant agreement in the JOLs associated with each item was found to decrease with repeated experience studying the same list. In addition, with repeated experience, the JOL–recall correlation increased with repeated presentations while the correlation between JOLs and judged item difficulty gradually decreased. Other studies have yielded a negative ST–JOL correlation for lists that consisted only of unrelated paired associates (Koriat, 2008; Koriat et al., 2009a). As we note later, the marked differences observed in the JOL results between Experiment 1 and Experiment 2 also weaken the argument that JOLs are based directly on judged item difficulty.

Assuming the distinction between data-driven and goal-driven regulation proposed by Koriat et al. (2006), the co-occurrence of both types of regulation within the same task should pose a challenge for the metacognitive monitoring of one's own learning. The observation of both a positive and a negative ST–JOL relationship was taken to suggest the operation of a delicate attribution process (Koriat & Nussinson, 2009). In this process, enhanced study effort is attributed to different degrees to data-driven and goal-driven variations, and metacognitive judgments either decrease or increase depending on that attribution. Our intention in this study was to trace the development of the ability underlying this attribution and to obtain some insight into its determinants and dynamics. We focused on children between the ages of 10 and 12 years; previous studies (Koriat et al., 2009a, 2009b) had suggested that at that age, children's JOLs are already sensitive to data-driven variation. We asked whether these children's JOLs also exhibit sensitivity to goal-driven variation and whether they also evidence, at the same time, sensitivity to data-driven variation. Let us first examine the pattern of results obtained for children at this age group before considering the metacognitive development that occurs later on.

Data-Driven and Goal-Driven Monitoring and Regulation in Primary School Children

The results of Experiments 1–3 provided a rather coherent picture. The children's metacognitive judgments exhibited sensitivity to both data-driven and goal-driven regulation but not in the same task. With regard to data-driven regulation, an inverse ST–

JOL relationship was observed in Experiment 2 and Experiment 3 (Phase 1), as well as in the constant-incentive condition of Experiment 3 (Phase 2). This relationship accords with the CM model of data-driven regulation.

With regard to goal-driven regulation, in both Experiment 1 and Experiment 3 (Phase 2), the fifth and sixth graders allocated different amounts of ST to different items in accordance with the incentive awarded for recall. Furthermore, there was a tendency for recall to increase with increasing ST, suggesting that children were somewhat successful in achieving differential memory performance for items that differed in their incentives (see Hanten et al., 2007). It is important to note that JOLs also increased with increasing ST, consistent with the MC model.

Most significant, however, is the observation that in none of the first three experiments did children exhibit *both* a negative and a positive ST–JOL relationship within the same task. Whereas Experiment 1 yielded evidence for metacognitive sensitivity to goal-driven variation in ST but not to data-driven variation, Experiment 2 indicated the opposite pattern. Experiment 3, in turn, in which we attempted to induce a more balanced sensitivity to both types of variation, yielded only sensitivity to data-driven variation. The results of Phase 2 of that experiment clearly confirmed that the same children exhibit sensitivity to either of the two types of variation when the salience of the variation along the other dimension was reduced.

Taken together, the results suggest that children at the ages of 10–12 are not capable of considering simultaneously the implications of data-driven and goal-driven variation. For example, introducing differential incentives into focus in Experiment 2 seemed to enhance sensitivity to goal-driven variation while impeding sensitivity to data-driven variation. It should be stressed that the inability to respond to both types of variation is not due to a general inability to consider the effects of two variables simultaneously. Wellman et al. (1981), for example, who had participants predict recall under hypothetical situations that varied in both number of items presented and degree of applied effort, found that even kindergarteners were able to take account of both variables in their predictions. Rather, the difficulty encountered by the children in this study seems to derive from the contrasting implications for JOLs of data-driven and goal-driven variation. As indicated earlier (see also Koriat et al., 2006), the pattern of results exhibited by college students implies a delicate factoring of ST variation between two potential sources and the derivation of an overall metacognitive judgment on the basis of that factoring. Presumably, the mechanism underlying children's metacognitive judgments is not sophisticated enough to allow this type of factoring and attribution. Let us examine these developmental changes more closely.

The Development of Data-Driven and Goal-Driven Regulation and Their Impact on Metacognitive Judgments

A comparison of the results for the three age groups—fifth and sixth graders, ninth graders, and college students—indicates an age-related increase in the ability to respond to both data-driven and goal-driven variation within the same task. We should note, however, that there were signs in our results for the burgeoning of this ability even among the youngest group. In Experiment 1, in which the dominant effects on JOLs were those of goal-driven regulation, there was also a trend indicating higher JOLs for short-ST items than for long-ST items. In parallel, Experiment 2 and Experiment 3 (Phase 1), which yielded significant effects on JOLs only for data-driven variation, also indicated a trend suggesting longer STs and higher JOLs for 5-point items than for 1-point items. Nevertheless, the overall sensitivity to *both* data-driven and goal-driven variation clearly increased with age. This is suggested by the sum of the effect sizes for data-driven and goal-driven regulation to JOLs (Table 1) and by the amount of variance in JOLs that is accounted for by both types of variation. In comparison to ninth graders (ages 14–15 years), the fifth and sixth graders (ages 10–12 years) exhibited a deficiency in the ability to respond simultaneously to data-driven and goal-driven variation. Experiment 5 helped relate this deficiency to the inability to partition the ST variation between its two hypothesized sources. ST regulation under the partitioning procedure (Figure 7, Panel B) revealed the expected pattern of ST increasing with both item relatedness and incentive. In parallel, both JOLs and recall yielded an adult-like pattern, decreasing with data-driven variation and increasing with goal-driven variation.

On the whole, the findings point to developmental changes in monitoring that occur well beyond the primary school years, which is unlike what is implied by previous findings (Roebers et al., 2007). In fact, the results obtained for primary school children underscore the impressive, nontrivial achievement demonstrated even by ninth graders. First, the ninth graders' regulation of ST was tuned to both data-driven and goal-driven regulation. Second, their JOLs were tuned differentially to differences in ST according to their source. Finally, the differential sensitivity of JOLs was successful in mirroring faithfully the pattern obtained for actual recall (Figure 6).

Results reported by Koriat and Ma'ayan (2005) suggested that JOLs are based on the flexible and adaptive utilization of different mnemonic cues—encoding fluency and retrieval fluency—according to their relative validity in predicting memory performance. In the present study, however, we showed that this is true even when the same cue (ST, study effort) has contrasting cue validities depending on it source.

A question of interest is whether the partitioning procedure we used with the younger children simulates the process underlying the metacognitive regulation and monitoring that was exhibited by ninth graders and college students. The partitioning procedure entailed a temporal spacing of the data-driven and goal-driven contributions to ST. It is unclear how the younger children took advantage of that spacing to separate between the two types of contribution. One possibility is that they first applied a memorizing effort heuristic to reach a tentative JOL prior to the announcement of the incentive associated with the item. They then updated that JOL to take into account the extra effort invested in a goal-driven manner in response to the incentive announced. Indeed, the results from Phase 2 of Experiment 3 indicated that fifth and sixth graders are able to respond to both data-driven variation and goal-driven variation when variation along the other dimension is minimized. Thus, in the partitioning condition of Experiment 5, children could presumably consider each source of variation in turn, but they had also to combine the implications of the two variations in order to reach an overall JOL.

This analysis implies that children's behavior in Experiment 5 can be described in terms of a sequence in which the CM model is followed by the MC model. The first phase entails a bottom–up process in which ST is dictated by the item–learner interaction. The second phase, in contrast, involves greater executive control: More effort is invested until a desired overall JOL is reached that fits the announced incentive. In this second phase, the data-driven effort already invested in the first phase must be taken into account, and indeed when the incentive was low, children tended simply to move to the next item.

Assuming that this analysis captures the process underlying children's behavior in the partitioning condition, does it also describe the behavior of the older children and adults when the incentive is announced at the beginning of a trial? In principle, participants could put aside the incentive value and consider it only when the data-driven regulation ends. We suspect, however, that participants cannot suspend consideration of the incentive until a later stage. Clearly, more research is needed to clarify the details of the process underlying older participants' metacognitive monitoring when data-driven and goal-driven variations must be taken into account in tandem.

The Pliability of Metacognitive Monitoring and Regulation

Experiment 1 and Experiment 2 differed only in how the related and unrelated pairs were distributed across the two blocks of the experiment. Experiment 1 put an emphasis on goal-driven variation by slating the related and unrelated pairs to different blocks and manipulating incentives within blocks, whereas Experiment 2 emphasized data-driven variation by manipulating item relatedness within blocks. Yet, the patterns of JOL effects differed markedly between the two experiments, with the former yielding sensitivity only to goal-driven variation and the latter evidencing sensitivity only to data-driven variation. This difference testifies to the remarkable pliability of metacognitive judgments. JOLs are not only relative and comparative in nature (Koriat, 1997) but are highly sensitive to the local context in which they are made. The adaptability of metacognitive judgments to the context of learning has been emphasized by Osman and Stavy (2006). They proposed that the more an item differs from other stimuli in the task, the more salient this difference becomes, and the salient feature may then be used by people, even by children, in their reasoning process.

It is interesting to note the parallels between metacognitive regulation and metacognitive monitoring. In Experiment 1, learners' allocation of ST was more tuned to goal-driven variation, and in parallel, JOLs were also more sensitive to goal-driven variation. In contrast, in Experiment 2, STs were more tuned to data-driven variation, and JOLs were also more sensitive to that variation. This pattern suggests that JOLs tend to be responsive to the major dimension of ST variation.

The results just discussed also illustrate how minor procedural variations in the context of learning can produce dissociations between metacognitive monitoring and memory performance. In both Experiment 1 and Experiment 2, recall performance evidenced strong effects of data-driven variation: The difference between mean recall for short ST and long ST amounted to 13.75 in Experiment 1 and 8.13 in Experiment 2. In contrast, the results for JOLs differed markedly between the two experiments: The respective difference in JOLs between short ST and long ST was 1.68 in Experiment 1, but 9.20 in Experiment 2. Experiment 3 also yielded similar data-driven effects on recall (a 13.12 difference between short ST and long ST) but a more moderate effect on JOLs (a 6.16 difference). Thus, procedural changes can produce differences in JOLs that are not mirrored by recall performance. Such dissociations highlight the conditions in which the results for JOLs mirror faithfully the pattern observed for recall, as demonstrated by ninth graders (Figure 6).

Back to William James

We conclude this article by coming back to the metatheoretical issue raised by William James (1884) regarding the cause-and-effect relation between emotional experience and behavior (see Koriat et al., 2006). As noted earlier, the MC model, which is the dominant model in metacognition (see Koriat & Goldsmith, 1996; Metcalfe & Finn, 2008; Nelson & Narens, 1990; Son & Schwartz, 2002), is consistent with the view that subjective feelings drive behavior, whereas the CM model is consistent with the view that subjective feelings are based on the feedback from one's own behavior (Kelley & Jacoby, 1998; Niedenthal, 2007; Strack & Deutsch, 2004). Koriat et al. (2006) took their results to imply that the two models are not mutually exclusive and that metacognitive feelings (e.g., JOLs, subjective confidence) can drive control operations (e.g., ST allocation, response latency), but they can also be based on the feedback from control operations.

The results obtained in the present study for primary school children suggest that children behave as if the two models are indeed mutually exclusive, because they show evidence either for one model or for the other depending on the situation. Of course, the observation that they do exhibit evidence consistent with both models, albeit each in a different condition, argues for the importance of considering both models in analyzing young children's metacognitive processes. However, the results observed for ninth graders reinforce the importance of considering both models, because they yielded evidence for both models within the same situation. It should be noted that the two models can also occur successively, with monitoring being based on the feedback from control operations (CM) but that monitoring can then drive a new control operation (MC). Results reported by Koriat and Ackerman (2010) for confidence judgments suggest that this kind of concatenated CM–MC chain can be demonstrated even by primary school children. As noted earlier, even the partitioning condition of Experiment 5 may have involved CM–MC sequence.

References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied, 17,* 18–32. doi:10.1037/a0022086

Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency–confidence association in problem solving. *Psychonomic Bulletin & Review, 19,* 1187–1192.

Annevirta, T., & Vauras, M. (2001). Metacognitive knowledge in primary grades: A longitudinal study. *European Journal of Psychology of Education, 16,* 257–282. doi:10.1007/BF03173029

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

Baumeister, R. F., & Vohs, K. D. (2004). *Handbook of self-regulation: Research, theory, and applications*. New York, NY: Guilford Press.

Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology, 24,* 654–663. doi:10.1037/0012-1649.24.5.654

Castel, A. D., Humphreys, K. L., Lee, S. S., Galván, A., Balota, D. A., & McCabe, D. P. (2011). The development of memory efficiency and value-directed remembering across the lifespan: A cross-sectional study of memory and selectivity. *Developmental Psychology, 47,* 1553–1564. doi:10.1037/a0025623

Castel, A. D., Lee, S. S., Humphreys, K. L., & Moore, A. N. (2011). Memory capacity, selective control, and value-directed remembering in children with and without attention-deficit/hyperactivity disorder. *Neuropsychology, 25,* 15–24. doi:10.1037/a0020298

Demmrich, A. (2005). *Improving reading comprehension by enhancing metacognitive competences: An evaluation of the reciprocal teaching method* [Doctoral dissertation]. Universität Potsdam, Potsdam, Germany.

Derryberry, D., & Rothbart, M. K. (1997). Reactive and effortful processes in the organization of temperament. *Development and Psychopathology, 9,* 633–652. doi:10.1017/S0954579497001375

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: Erlbaum.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica, 98,* 37–56. doi:10.1016/S0001-6918(97)00051-6

Eisenberg, N., & Morris, A. S. (2002). Children's emotion-related regulation. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 30, pp. 189–229). Amsterdam, the Netherlands: Academic Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34,* 906–911. doi:10.1037/0003-066X.34.10.906

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50,* 21–45. doi:10.1146/annurev.psych.50.1.21

Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3–33). Hillsdale, NJ: Erlbaum.

Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *Quarterly Journal of Experimental Psychology, 60,* 991–1004. doi:10.1080/17470210600823595

Ghetti, S. (2008). Rejection of false events in childhood: A metamemory account. *Current Directions in Psychological Science, 17,* 16–20. doi:10.1111/j.1467-8721.2008.00540.x

Hanten, G., Chapman, S. B., Gamino, J. F., Zhang, L., Benton, S. B., Stallings-Roberson, G., . . . Levin, H. S. (2004). Verbal selective learning after traumatic brain injury in children. *Annals of Neurology, 56,* 847–853. doi:10.1002/ana.20298

Hanten, G., Li, Z., Chapman, S. B., Swank, P., Gamino, J., Roberson, G., & Levin, H. S. (2007). Development of verbal selective learning. *Developmental Neuropsychology, 32,* 585–596. doi:10.1080/87565640701361112

Hanten, G., Zhang, L., & Levin, H. S. (2002). Selective learning children after traumatic brain injury: A preliminary study. *Child Neuropsychology, 8,* 107–120. doi:10.1076/chin.8.2.107.8729

Hoffmann-Biencourt, A., Lockl, K., Schneider, W., Ackerman, R., & Koriat, A. (2010). Self-paced study time as a cue for recall predictions across school age. *British Journal of Developmental Psychology, 28,* 767–784. doi:10.1348/026151009X479042

James, W. (1884). What is an emotion? *Mind, 9,* 188–205. doi:10.1093/mind/os-IX.34.188

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

Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica, 98,* 127–140. doi:10.1016/S0001-6918(97)00039-5

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. doi:10.1006/jmla.1993.1001

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling-of- knowing. *Psychological Review, 100,* 609–639. doi:10.1037/0033-295X.100.4.609

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. doi:10.1037/0096-3445.126.4.349

Koriat, A. (2006). Are we frightened because we run away? Some evidence from metacognitive feelings. In B. Uttl, N. Ohta, & A. L. Siegenthaler (Eds.), *Memory and emotion: Interdisciplinary perspectives* (pp. 83–103). Malden, MA: Blackwell.

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition, 36,* 416–428. doi:10.3758/MC.36.2.416

Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science, 13,* 441–453. doi:10.1111/j.1467-7687.2009.00907.x

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009a). The easily learned, easily remembered heuristic in children. *Cognitive Development, 24,* 169–182. doi:10.1016/j.cogdev.2009.01.001

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009b). The memorizing-effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology, 102,* 265–279. doi:10.1016/j.jecp.2008.10.005

Koriat, A., & Bjork, R. A. (2006a). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition, 34,* 959–972. doi:10.3758/BF03193244

Koriat, A., & Bjork, R. B. (2006b). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1133–1145. doi:10.1037/0278-7393.32.5.1133

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517. doi:10.1037/0033-295X.103.3.490

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. doi:10.1016/j.jml.2005.01.001

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. doi:10.1037/0096-3445.135.1.36

Koriat, A., & Nussinson, R. (2009). Attributing study effort to data-driven and goal-driven effects: Implications for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1338–1343. doi:10.1037/a0016374

Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy: Insights from the processes underlying judgments of learning in children. In P. Chambres, M. Izaute, & P.-J. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 1–17). New York, NY: Kluwer. doi:10.1007/978-1-4615-1099-4_1

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 609–622. doi:10.1037/0278-7393.32.3.609

Krebs, S. S., & Roebers, C. M. (2012). The impact of retrieval processes, age, general achievement level, and test scoring scheme for children's metacognitive monitoring and controlling. *Metacognition & Learning, 7,* 75–90. doi:10.1007/s11409-011-9079-3

Kreutzer, M. A., Leonard, C., & Flavell, J. H. (1975). An interview study of children's knowledge about memory. *Monographs of the Society for Research in Child Development, 40,* 1–60. doi:10.2307/1165955

Kunzinger, E. L., & Witryol, S. L. (1984). The effects of differential incentives on second-grade rehearsal and free recall. *Journal of Genetic Psychology, 144,* 19–30. doi:10.1080/00221325.1984.10532447

Lockl, K., & Schneider, W. (2002). Developmental trends in children's feeling-of-knowing judgments. *International Journal of Behavioral Development, 26,* 327–333. doi:10.1080/01650250143000210

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. doi:10.1080/17405620444000085

Lyons, K., & Ghetti, S. (2010). Metacognitive development in early childhood: New research challenges and old assumptions. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 259–278). Hauppauge, NY: Nova. doi:10.1007/978-1-4419-6546-2_12

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15,* 174–179. doi:10.3758/PBR.15.1.174

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (in press). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133. doi:10.1037/0033-2909.95.1.109

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5,* 207–213. doi:10.1111/j.1467-9280.1994.tb00502.x

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. doi:10.1037/0278-7393.14.4.676

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, NY: Academic Press.

Niedenthal, P. M. (2007, May 18). Embodying emotion. *Science, 316,* 1002–1005. doi:10.1126/science.1136930

Ojose, B. (2008). Applying Piaget's theory of cognitive development to mathematics instruction. *The Mathematics Educator, 18,* 26–30.

Osman, M., & Stavy, R. (2006). Development of intuitive rules: Evaluating the application of the dual-system framework to understanding children's intuitive reasoning. *Psychonomic Bulletin & Review, 13,* 935–953. doi:10.3758/BF03213907

O'Sullivan, J. T. (1993). Preschooler's beliefs about effort, incentives, and recall. *Journal of Experimental Child Psychology, 55,* 396–414. doi:10.1006/jecp.1993.1022

O'Sullivan, J. T. (1997). Effort, interest, and recall: Beliefs and behaviors of preschoolers. *Journal of Experimental Child Psychology, 65,* 43–67. doi:10.1006/jecp.1996.2355

Piaget, J. (1977). The role of action in the development of thinking. In W. F. Overton & J. M. Gallagher (Eds.), *Knowledge and development* (pp. 17–42). New York, NY: Plenum Press.

Roderer, T., & Roebers, C. M. (2008). Children's strategic regulation of memory accuracy. In M. R. Kelley (Ed.), *Applied memory* (pp. 253–274). Hauppage, NY: Nova Science.

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. doi:10.1016/j.jecp.2006.12.006

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. doi:10.1111/j.1751-228X.2008.00041.x

Schneider, W. (2010). The development of metacognitive competencies. In B. Glatzeder, V. Goel, & A. von Müller (Eds.), *Towards a theory of thinking: Building blocks for a conceptual framework* (pp. 203–214). Berlin, Germany: Springer.

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*. Hillsdale, NJ: Erlbaum.

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. doi:10.1037/0278-7393.26.1.204

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, England: Cambridge University Press.

Stipek, D. J., Roberts, M. T., & Sanborn, M. E. (1984). Pre-school age children's performance expectations for themselves and another child as a function of the incentive value of success and the salience of past performance. *Child Development, 55,* 1983–1989. doi:10.2307/1129773.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220–247. doi:10.1207/s15327957pspr0803_1

Strack, F., & Neumann, R. (2000). Furrowing the brow may undermine perceived fame: The role of facial feedback in judgments of celebrity. *Personality and Social Psychology Bulletin, 26,* 762–768. doi:10.1177/0146167200269002

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. doi:10.1037/0022-0663.95.1.66

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. doi:10.1037/0278-7393.25.4.1024

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language, 64,* 109–118. doi:10.1016/j.jml.2010.11.002

Wellman, H. M., Collins, J., & Glieberman, J. (1981). Understanding the combination of memory variables: Developing conceptions of memory limitations. *Child Development, 52,* 1313–1317. doi:10.2307/1129522

Zacks, R. T. (1969). Invariance of total learning time under different conditions of practice. *Journal of Experimental Psychology, 82,* 441–447. doi:10.1037/h0028369

Received August 6, 2012
Revision received January 2, 2013
Accepted January 3, 2013 ■