Examining on-task regulation in school children: Interrelations between monitoring, regulation, and task performance

van Loon, Mariëtte H ; Oeri, Niamh S

DOI: https://doi.org/10.1037/edu0000781

Originally published at:
van Loon, Mariëtte H; Oeri, Niamh S (2023). Examining on-task regulation in school children: Interrelations between monitoring, regulation, and task performance. Journal of Educational Psychology, 115(3):446-459.
DOI: https://doi.org/10.1037/edu0000781

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

2023, Vol. 115, No. 3, 446–459
https://doi.org/10.1037/edu0000781

Examining On-Task Regulation in School Children: Interrelations Between Monitoring, Regulation, and Task Performance

Mariëtte H. van Loon and Niamh S. Oeri
Institute for Psychology, University of Bern

It is unknown how multiple components of on-task regulation of learning affect task performance in school children. This research aimed to acquire insights into the interrelations between children's metacognitive monitoring, regulation of learning, and task performance. Three components of on-task regulation of learning were investigated: allocation of study time, restudy selections, and task persistence. Children learned concepts with their definitions. In Study 1, 104 sixth graders (M_{age} 12 years) participated; Study 2 consisted of 97 fourth graders (M_{age} 10 years). For both age groups, task persistence was a strong predictor of performance. For sixth but not for fourth graders, monitoring accuracy affected performance. Findings indicate that, when aiming to improve regulation of learning and task performance in elementary school, student age is a relevant factor to consider. Around the age of 10, regulation affects learning performance, whereas the effects of self-monitoring accuracy on performance seem apparent when children are approximately 12 years of age.

Educational Impact and Implications Statement
Student age is a critical factor when aiming to improve the regulation of learning and performance. For fourth graders, teachers may address how on-task regulation can be effective and particularly focus on task persistence. For sixth graders, monitoring accuracy was a strong predictor of performance. Around the age of 12, addressing and training self-monitoring skills could play a prominent role during classroom teaching.

Keywords: late childhood, restudy, self-monitoring, study time allocation, task persistence

Children need to regulate their learning in multiple ways in school. For instance, they need to sustain learning efforts, persist when difficulties arise, and restudy materials when identified as not yet understood (Boekaerts & Corno, 2005; Pintrich, 2002). Children's regulation predicts their learning performance (e.g., Cleary & Zimmerman, 2004; Ladd & Dinella, 2009; Lawson & Lawson, 2013). Therefore, it is important to understand the processes at play when children regulate their learning. These insights would seem particularly relevant for children in late childhood (approximately age 10–12), who are preparing for the transition from elementary to middle schools or secondary schools, where they have to take more responsibility for their learning. However, children's regulation processes have mainly been investigated using questionnaires; the topic of children's regulation *during* work on tasks has received less attention. When aiming to teach children how to regulate learning effectively, it is crucial to know which regulation processes are most predictive of performance; this study aims to acquire insights into this.

There is consensus that regulation of learning consists of several subprocesses (Ackerman & Goldsmith, 2011; Callan & Cleary, 2018). Although a few studies measured children's on-task regulation during learning, these addressed only one subprocess: restudy selections (Dufresne & Kobasigawa, 1989) or study time allocation (Metcalfe & Finn, 2013). It remains unknown to what extent multiple components of on-task regulation relate to each other and how regulation components predict children's learning. The present research addresses these issues by measuring three on-task subprocesses of regulation of learning: (a) study time allocation, (b) restudy selections, and (c) task persistence.

The assumption that self-monitoring, regulation, and performance are related is based on the theoretical framework of metacognition by Nelson and Narens (1990). This framework describes learning processes through two interrelated levels, the meta-level and the object-level, and proposes a direction for the information flow between these levels. The meta-level is informed by the object-level through monitoring. Based on monitoring, the meta-level then modifies the object-level through regulation. Most importantly, monitoring influences regulation decisions, which then affect performance (i.e., the object level). Accurate self-monitoring is thus presumed to be a prerequisite for adequate regulation and high task performance

This article was published Online First January 9, 2023.
Niamh S. Oeri https://orcid.org/0000-0003-3685-8220
All data and research materials are available at https://osf.io/y25zk/?view_only=b5e2a264a47544119aa9b1f4cbf10e72. The authors would like to thank Janet Metcalfe and Lisa Son for their comments on a previous draft of this manuscript.

Mariëtte H. van Loon served as lead for investigation, methodology, project administration, and software. Mariëtte H. van Loon and Niamh S. Oeri contributed to formal analysis, writing–original draft, and writing–review and editing equally.

Correspondence concerning this article should be addressed to Mariëtte H. van Loon, Institute for Psychology, University of Bern, Fabrikstrasse 8, 3006 Bern, Switzerland. Email: mariette.vanloon@unibe.ch

(De Bruin & Van Gog, 2012; Nelson & Narens, 1990, 1994). Through self-monitoring, a person is aware of the discrepancy between goals and the present state of learning. In response to self-monitoring, learning actions can be taken to reach learning goals (Moilanen, 2007).

The COPES model of self-regulated learning extends the model by Nelson and Narens (1990) and offers a further framework for research on the interactive nature between monitoring, regulation, and performance during work on learning tasks (Winne & Hadwin, 1998, 2008). The acronym stands for Conditions, Operations, Products, Evaluations, and Standards. Monitoring (referred to as evaluations) and regulation (referred to as operations) are central features of this model which are connected to performance changes (referred to as products in the COPES model or the object-level in the Nelson & Narens model). Through evaluation, a learner monitors whether there is a discrepancy between the actual products and the learning standards. Based on these evaluations, a learner then operates to change outcomes. Notably, the COPES model emphasizes the flexible and recursive on-task processes of self-monitoring and regulation and performance. While working on a task, monitoring informs regulation, which affects learning products (i.e., the object level). These changes in the learning products then inform subsequent monitoring and regulation.

Although many studies offer evidence that students can tell which actions they might undertake to control learning (Dent & Koenka, 2016), minimal research examines the actual, theoretically proposed relations between on-task monitoring, regulation, and learning performance (for an exception with adult participants, see Thiede et al., 2003). It remains largely unknown to what extent accurate self-monitoring benefits regulation and performance for children. Gaining insights into the relations between metacognitive monitoring, regulation, and task performance is highly relevant. When tailoring interventions to support children's learning, it is essential to know if the focus should lie on learning regulation or self-monitoring processes. Therefore, a further aim of the present research is to investigate to what extent children's regulation and task performance are affected by their self-monitoring judgments.

On-Task Regulation of Learning

Regulation of learning has mainly been investigated using self-report, parent-report, and teacher-report instruments (Dent & Koenka, 2016). Although self-report data provide insights into beliefs about how learning is regulated, such measurements may not necessarily capture the actual actions during real-time learning (Cromley & Azevedo, 2006). When aiming to understand how metacognitive regulation affects learning, it seems necessary to include regulation measures "in action" captured while working on a task.

Research used allocation of study time (e.g., Son & Kornell, 2008; Thiede & Dunlosky, 1999), restudy selections (e.g., Metcalfe & Finn, 2013), and task persistence (e.g., Andersson & Bergman, 2011) as on-task measures to indicate the effectiveness of regulation. Learning performance seems to benefit most when study time is strategically allocated to task items that are experienced as difficult and not yet well learned (Dufresne & Kobasigawa, 1989; Hines et al., 2009; Son & Kornell, 2008). Further, regulation is considered effective when not well-learned materials are selected for restudy (e.g., Metcalfe & Finn, 2013; Thiede et al., 2017; Van Loon et al., 2013).

Notably, the dynamic nature of regulation of learning can only be investigated when persons have complete responsibility for their learning and can quit the task whenever they like to (Dignath & Büttner, 2008; Mokrova et al., 2013; Zhou et al., 2007). Task persistence also seems highly relevant as an indicator of regulation in the educational context. Higher task persistence is related to better school outcomes for children and adolescents, and longitudinally, children's task persistence seems to be related to educational attainment in adulthood (Andersson & Bergman, 2011). Persistence is defined as a person's tendency to endure and has been treated in some studies as a personal trait and in other studies as a state-level indicator showing to what extent a person proceeds toward a specific external or internal goal (Howard & Crayne, 2019). In the present research, task persistence is investigated at state-level to indicate to what extent a person reaches the goal of learning the task materials. Although in theory possible, the task is too difficult to complete entirely and to get fully correct performance (in line with tasks designed to measure persistence by Oeri & Roebers, 2021). The used measure of task persistence captures to what extent students stick to the task and at which point they decide to give up and stop the task.

A unique contribution of the present study is that multiple subprocesses of on-task regulation of learning (i.e., study time allocation, restudy, and task persistence) are investigated in conjunction. Firstly, this study addresses the relations between these components of regulation. Moreover, this research aims to acquire insights into the extent to which these processes of on-task regulation affect children's performance.

Effects of Monitoring Accuracy on Regulation

Presumably, self-regulated learning can only be effective when self-monitoring is accurate (De Bruin & Van Gog, 2012; Nelson & Narens, 1990, 1994). In part, these assumptions are justified by research findings. Children's accurate self-monitoring seems to benefit the effectiveness of their study time allocation and restudy selections (Schneider & Löffler, 2016). For instance, from the age of 6 onwards, children allocate more study time to items for which they give low monitoring judgments than items for which judgments are high (Destan et al., 2014). Furthermore, from the age of 8 onwards, children's monitoring seems related to restudy selections, such that task items for which judgments are low are more often selected than items for which judgments are high (Van Loon et al., 2017).

Relations between self-monitoring and task persistence seem less clear; there may even be advantages of inaccurate monitoring for task persistence. A study by Shin et al. (2007) showed that 6–9-year-old children who overestimated their learning had higher recall performance than children who accurately monitored performance. Based on these findings, Shin et al. suggested that overconfidence may benefit children's task persistence and therefore improve learning. However, children could not decide how long they liked to persist with the task before giving up. Thus, the assumption that overconfidence benefits task persistence was not explicitly tested.

Note that in research on monitoring accuracy, two different analytical approaches are used to investigate relations between monitoring judgments and task performance (Dunlosky et al., 2016; Dunlosky & Thiede, 2013). Relative accuracy measures show the strength of the relation between monitoring judgments and performance across task items (typically measured with intra-individual correlations between judgments and performance). Measures of absolute monitoring accuracy indicate how magnitudes of monitoring judgments match performance measures and show to what extent persons are accurately calibrated, overconfident, or underconfident. In the present research, we use measures of absolute monitoring accuracy. Specifically, we address to what extent overconfidence is beneficial (as suggested by Shin et al., 2007) or harmful for regulation and performance (as suggested by Dunlosky & Rawson, 2012; Rinne & Mazzocco, 2014).

Interrelations Between Monitoring, Regulation, and Learning Performance

According to the model by Nelson and Narens (1990), monitoring, regulation, and learning performance are intertwined (for evidence supporting this assumption see De Bruin & Van Gog, 2012; Dunlosky & Rawson, 2012; Rinne & Mazzocco, 2014; Winne & Hadwin, 2008). However, to the best of our knowledge, for children, only one study (Metcalfe & Finn, 2013) measured monitoring, on-task regulation, and performance. In this study, on-task-regulation was investigated by giving children the opportunity to restudy items. Findings showed that relations between regulation and performance might be age-dependent. Ten-year-old children strategically restudied items for which monitoring judgments were low, whereas, for 8-year-old, there were no relations between monitoring and restudy. Moreover, findings did not support assumptions on the intertwined relations between monitoring, regulation, and performance. That is, even for 10-year-old children, accurate monitoring and subsequent effective restudy did not improve recall performance. However, in Metcalfe and Finn (2013), the tasks were experimenter-paced; children did not have the autonomy to decide how much study time they allocated and how long they liked to persist before stopping. Therefore, the interrelations between monitoring, regulation, and performance may have been underestimated. Thus, for the present research, a task was developed for which children had the autonomy to regulate their learning flexibly. That is, they could decide how long they study, what they restudy, and how long they persist before giving up.

Present Study

The present study aims to (a) investigate relations between components of children's (late childhood, ages 10–12) on-task regulation of learning; (b) analyze whether children's monitoring judgments affect regulation; and (c) investigate the extent to which monitoring accuracy and the effectiveness of regulation predict children's task performance. Children worked on a concept task for which they studied the definitions of difficult concepts and were then tested for memory of the concept definitions. Three regulation components were measured: allocation of study time, restudy selections, and task persistence.

Research using self-report scales showed interrelations between different components of the regulation (Callan & Cleary, 2018; Pintrich et al., 1994). Based on these findings, we assumed that also when using on-task measures, the three components of on-task regulation would be related to one another (Hypothesis 1). Moreover, we expected that monitoring judgments affect the subsequent allocation of study time (cf. Destan et al., 2014; Hypothesis 2a) and restudy selections (cf. Van Loon et al., 2017; Hypothesis 2b). Findings on the relations between self-monitoring and task persistence are mixed: according to Shin et al. (2007) overconfidence may be beneficial whereas Dunlosky and Rawson (2012) and Rinne and Mazzocco (2014) suggest disadvantageous effects. Therefore, no hypothesis was formulated about this relation. Furthermore, although the Nelson and Narens model (1990) assumes interrelations between monitoring, regulation, and performance, findings are mixed for children. Therefore, we did not formulate specific hypotheses about the extent to which monitoring accuracy and components of regulation would predict children's task performance.

The present research consists of two separate studies. In Study 1, participants were 12-year-old sixth-grade students. To foreshadow, we did find evidence for relations between monitoring, regulation, and learning. To investigate whether these findings could also be replicated for a younger age group, in Study 2, 10-year-old fourth graders participated.

Study 1

Participants

Participants were 104 sixth graders (M_{age} 12.33 years, *SD* 5.2 months, 54 females) attending public schools in the German-speaking part of Switzerland. All participants were used to following school instructions in the German language. Informed consent was received from parents/caregivers. Children were told that they could withdraw at any time without consequences. Participants were tested in the context of a more extensive study, for which they also completed questionnaires on self-efficacy and learning goals before starting the concept learning task. Due to the scope of this article on investigating interrelations between monitoring, regulation, and task performance, only the findings for the concept learning task are reported.

Materials and Procedure

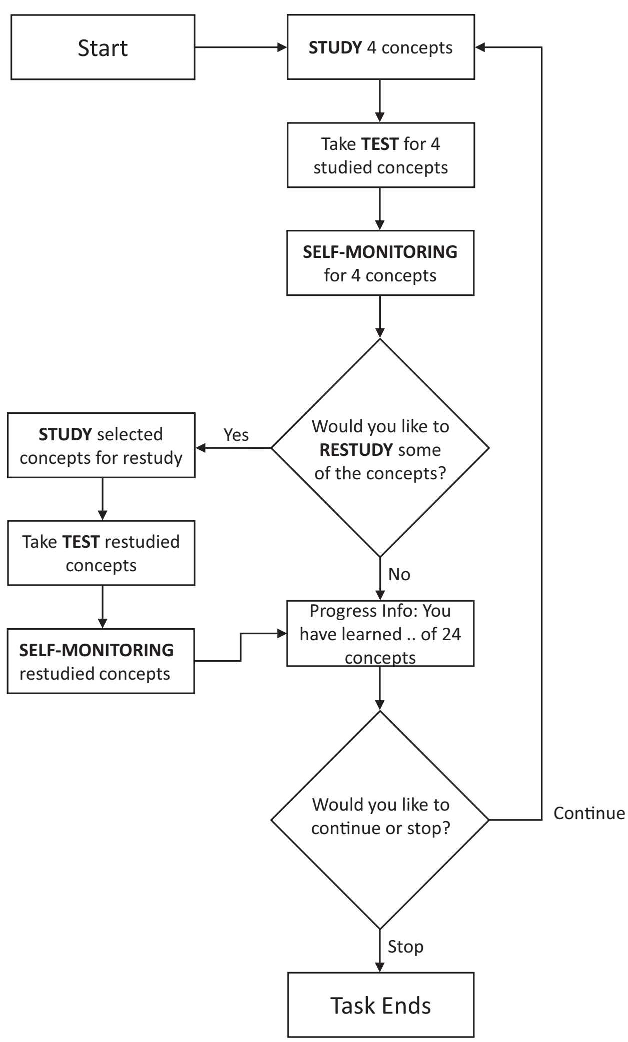
Children were tested in their classrooms. For each testing session, 90 min were reserved so that they would have sufficient time available to allocate to the task. Depending on class size, two to four researcher assistants visited the class. After the session, all children received a small thank-you gift.

For the concept learning task, children could learn a maximum of 24 difficult concepts with definitions. The concepts were initially developed for studies by Van Loon and Roebers (2017) and Van Loon and Van de Pol (2019). These concepts were based on teaching materials and the learning objectives outlined in the national school curriculum (Lehrplan 21, www.lehrplan21.ch). The difficult concepts were taken from school materials in geography, history, biology, physics, chemistry, and social sciences. To ensure low prior knowledge and a sufficient difficulty level, concepts were selected that were to be learned in the following school years.

Concept definitions consisted of three idea units. In addition to the definition, an example sentence was presented in which the concept was used. According to Rawson et al. (2015) and Zamary and Rawson (2018) , examples may be effective as supplementary support when learning abstract concepts. Therefore, the example sentence was included to clarify the meaning of the concept definitions. For instance, the concept Phobia was presented as follows: Phobia (concept)—A Phobia is in which one feels very great fear for which there is no real reason (definition)—Anna is extremely afraid of spiders and therefore suffers from a spider phobia (example sentence). The concept learning task was completed on a tablet equipped with a keyboard. A pilot study was conducted with eight sixth graders to ensure that children could deal with the digitalized concept learning task.

Concepts were grouped in blocks of four. For each child, concepts were randomly assigned to these blocks, so the order of concept presentation differed for every participant. Each block consisted of a study phase, test phase, self-monitoring phase, an optional restudy phase (with subsequent retest and monitoring), progress information, and a stop/continue decision. Figure 1 shows the task procedure. Before starting the task on the tablet, one of the researchers explained the task to the whole class. The children were shown a poster with two practice concepts with definitions and example sentences, examples of the test questions, and the monitoring scale. This poster explained that all concept definitions consisted of three idea units. Per test response, a child could get a score ranging from 0 (*no idea units from the definition are present*) to 3 (*all idea units from the definition are present*). Children were told that with their self-monitoring judgments, they would score their performance. More specifically, for each test response, children had to judge how accurately the three idea units were remembered. It was explained that children could study a maximum of 24 difficult concepts and that the task goal was to learn as many concept definitions as well as possible. Further, children were told that they could use the time they needed to study the concepts, had the option to restudy a concept once to improve their test response, and stop the task when they liked to.

Figure 1
Flowchart Depicting the Task Procedure



Study Phase

In each block, children studied the four concepts sequentially; each concept was presented on a screen with its definition and an example sentence. The study was self-paced.

Test and Self-Monitoring Phase

After learning a block of four concepts, the children took a test. They were asked to type the definition for each concept in a text field. Immediately after completing a test response, children monitored the quality of this response by making a self-score judgment (cf. Rawson & Dunlosky, 2007). With this judgment, they indicated how many idea units they thought they presented correctly in their response. To measure absolute monitoring accuracy of item-by-item judgments, monitoring judgments must be made on the same scale as used for scoring performance (Dunlosky et al., 2016). Therefore, the judgments were made by checking a box on a scale ranging from 0 (*no idea unit correct*) points to 3 (*all idea units correct*).

Restudy Phase

After making the monitoring judgment for the fourth tested concept (i.e., the last concept in the according block), children were asked if they would like to restudy any of the concepts. They could either click yes or no. If they clicked yes, the four concepts were presented simultaneously in a 2×2 grid (Thiede & Dunlosky, 1999, recommend this grid format); they could select 1–4 concepts for restudy. Then, the task continued with a restudy phase for the selected concepts. Identical to the first study phase, restudy of the selected concepts was self-paced. After restudy, children were again tested for the restudied concepts and made a self-score judgment. Note that children could only restudy the concepts once. As shown in Figure 1, if children did not select any concepts for restudy, the task continued with the next phase in which children were presented with progress information.

Continue/Stop Decision

Children were presented with a number line (ranging from 4 to 24 concepts), showing how much of the task they completed and how much was left. They were then asked whether they liked to continue or stop. If they decided to continue, the task continued with a study phase for a new block consisting of four novel concepts. If they decided to stop (or after studying the 24th and last concept), the task ended.

Statistical Analyses

Table 1 shows the measured variables indicating performance, monitoring, and regulation and describes how these were used for the analyses. For performance, test responses given after study trial 1 and after restudy were coded separately. Each test response was coded based on the number of idea units listed. For instance, the concept of Phobia was scored according to whether the following idea units were mentioned: (a) a disorder (b) in which one feels very great fear (c) for which there is no real reason. Similar to the scoring procedure by Rawson and Dunlosky (2007), the idea units were counted as being present in a response if they were stated verbatim or if they were a correct paraphrase of the original concept definition. A test response was assigned a score of zero if it contained no idea units and three if it contained all of the idea units from the definition.

The measure of total test performance indicates how much of the concepts children learned after completing study trial 1 as well as the optional restudy trial. When children improved their test responses after restudy, this performance score was used to measure final performance. For instance, if the score for a concept after study trial 1 was two, and then a person had a score of three after restudy, this improved score was used as an index of the final performance. Two independent raters double-coded 28% of the test responses; interrater reliability was high (ICC 0.92).

For monitoring accuracy, a calibration score of zero indicates perfect monitoring, deviations >0 show overconfidence, whereas deviations <0 indicate underconfidence. Measures of monitoring accuracy were calculated both for the judgments for test responses after study trial 1 and for judgments for test responses after restudy.

As outlined in Table 1, children's regulation was investigated with measures of study time allocation, restudy selections, and task persistence. Study time allocation was measured separately during study trial 1 and the restudy trial.

Note that the regulation measures and performance measures were right-skewed and not normally distributed. The right skew indicates that some children persisted much longer, effectively restudied many more items, allocated much more time when studying, and performed much better than most other children in the sample. With large enough sample sizes (>40), using parametric procedures to investigate sample means should not cause problems when the data are not normally distributed (Ghasemi & Zahediasl, 2012). Therefore, paired-sample *t* tests were used for preliminary analyses investigating performance and monitoring judgments before and after restudy.

When addressing our research questions and hypotheses on interrelations between monitoring, regulation, and performance, we used non-parametric measures to account for violations of the normality assumptions. More specifically, Spearman correlations were calculated to investigate relations between the following regulation components: allocation of study time, restudy selections, and task persistence (Hypothesis 1). To examine relations between monitoring and allocation of study time (Hypothesis 2a), intra-individual gamma correlations (as recommended by Nelson, 1984, to account for the ordinal measurement of self-score judgments) were calculated on item-level between self-score judgments after study trial 1 and allocation of study time during restudy. To investigate relations between monitoring and restudy selections (Hypothesis 2b), intra-individual gamma correlations between self-score judgments (per concept) and restudy selections were calculated. Because task persistence was measured on the task level, Spearman correlations were calculated to assess the relation between mean monitoring judgments and task persistence.

Furthermore, we computed path analyses (in R with the package Lavaan; Rosseel, 2012) to analyze the interrelations between monitoring accuracy, regulation (study time allocation, effective restudy selections, and task persistence), and performance. Two separate analyses were computed, one for mean performance per studied concept and one for total test performance. Error variances of the regulation components were allowed to correlate. The nonparametric bootstrapping procedure was used to account for the non-normal distributions in the data (1000 resamples; Preacher & Hayes, 2008).

Transparency and Openness

The ethical committee of the Faculty of Human Sciences, University of Bern, Switzerland approved the task materials and procedure. The design was not preregistered. As we did not have clear indications of potential effect sizes, we could not conduct a priori power analyses to determine the sample size. We anticipated that to be able to document medium effect sizes, we would need to have a sample of approximately 100 participants.

All data and research materials are openly available on the Open Science Framework OSF (Van Loon & Oeri, 2022). For the correlational analyses and *t* tests, data were analyzed using SPSS, Version 27. For the path analysis, data were analyzed with R and RStudio, Version 1.4.1717, and the R-package Lavaan (Rosseel, 2012).

Table 1
Research Variables

Variable	Description
Mean performance per concept	Mean test score for each child per learned concept (ranging from 0 to 3), indicating how many learned idea units were present in the test response.
Total test performance	Score for the final performance (which could range from 0 to 72), indicating how many idea units were correct for all responses combined, after finishing the task.
Self-score judgments	Monitoring judgments given for each test response, indicating the expected number of correct idea units (range 0–3).
Monitoring accuracy	Mean calibration score per child, indicating the mean deviance (which could range between −3 and +3) between self-score judgments and performance.
Study time allocation	Mean time each child allocated for study per learned concept, in seconds.
Restudy selections	Percentage of the learned items in study trial 1 that were subsequently selected for restudy.
Effective restudy selections	Percentage of restudy selections for the items that were incorrect at the test after study trial 1.
Task persistence	Percentage of studied items, consisting of studied items during study trial 1 (range 4–24 items) plus restudied items (which could range from 0 to 24 items).

Table 2
Performance and Self-Score Judgments

Grade	Performance per concept		Total task performance		Self-score judgment per concept	
	Study trial 1	Restudy trial	Study trial 1	Restudy trial	Study trial 1	Restudy trial
Six (Study 1)	1.37 (0.63)	1.57 (0.63)	22.45 (14.66)	25.62 (15.66)	1.79 (0.53)	2.13 (0.70)
Four (Study 2)	0.99 (0.60)	1.15 (0.68)	13.23 (11.02)	15.30 (12.42)	1.47 (0.66)	1.75 (1.02)

Note. Mean values are shown for both age groups, standard deviations in parentheses.

Results

Preliminary Analyses

Table 2 shows descriptive statistics for performance measures after study trial 1 and after restudy. Performance improved through restudy, $t(105) = 9.21$, $p < .001$.

Further, Table 2 shows the magnitude of self-score judgments after study trial 1 and for restudied concepts. After study trial 1, monitoring accuracy (as indicated with the calibration score) was 0.42 ($SD =$ 0.42); this value was higher than zero, indicating overconfidence, $t(105) = 10.26$, $p < .001$. For restudied items, calibration was 0.38 ($SD = 0.67$), and children were overconfident, $t(77) = 4.98$, $p < .001$. There was no difference in monitoring accuracy between judgments after study trial 1 and after restudy, $t(77) = 0.21$, $p = .83$.

Table 3 shows descriptive statistics for the allocation of study time, restudy selections, effective restudy, and task persistence. Note that 26.4% of the children did not select any concepts for restudy. Task persistence ranged from 8.33% to 83.33%. In addition, for a more detailed overview, Table 4 shows monitoring, regulation, and performance across the subsequent task blocks.

Interrelations Between Monitoring, Regulation, and Performance

Correlations Between Regulation Components

Table 5 presents the Spearman correlations between the regulation components. Measures of study time allocation were not related to restudy selections. Moreover, allocation of study time during trial 1 and overall allocation of study time were negatively related to measures of task persistence. That is, children who allocated more study time per concept got less far into the task (i.e., studied fewer concepts) than children who allocated less study time. There was a positive but weak relationship between effective restudy and task persistence. This indicates that children who more effectively restudied not-yet-learned items were also more likely to get further into the task. In sum, expectations that relations between regulation components would be positive (Hypothesis 1) were only partially confirmed.

Relations Between Self-Monitoring and Regulation

Contrary to Hypothesis 2a, there was no statistically detectable relation between monitoring judgments for study trial 1 and subsequent allocation of study time during restudy, Gamma $= -.03$, $SD = 0.75$, $t(47) = -0.29$, $p = .77$. Confirming Hypothesis 2b, monitoring judgments were related to restudy selections, Gamma $= -.73$, $SD = 0.46$, $t(77) = -14.04$, $p < .001$. That is, children mainly selected concepts for restudy for which their monitoring judgments were low.

There was no relation between self-monitoring and total task persistence (Spearman's $\rho = .037$, $p = .71$). However, possibly not the magnitude of monitoring judgments, but particularly overconfidence may be related to higher task persistence (Shin et al., 2007). To assess this, the correlation between monitoring and task persistence was calculated for the 86 children who showed overconfidence (i.e., had a calibration value higher than zero). There was no relation between the degree of overconfidence and total task persistence (Spearman's $\rho = -.046$, $p = .68$).

Effects of Monitoring and Regulation on Performance

Standardized coefficients from the path analysis for mean performance per concept are presented in Figure 2 and Table 6 shows the unstandardized coefficients and the 95% confidence intervals. The effect of monitoring accuracy on performance indicates that less overconfident children had higher performance. Self-monitoring accuracy was the strongest predictor of mean performance per concept.

Further, monitoring accuracy affected study time allocation (such that more accurate monitoring was related to allocating more time per concept) and effective restudy selections (such that more accurate monitoring was related to improved restudy of initially incorrect responses). Both study time allocation and effective restudy were predictive of mean performance per concept.

Figure 3 shows the path analysis coefficients (standardized) for the effects of monitoring accuracy and regulation on total task performance; unstandardized coefficients and 95% confidence intervals are presented in Table 7. Monitoring accuracy predicted study time

Table 3
Components of Regulation: Study Time Allocation, Restudy Selections, and Task Persistence

Grade	Study time allocation			Restudy		Task persistence	
	Trial 1	Restudy trial	Restudy selections	Effective restudy selections	Concepts studied trial 1	Concepts restudied	Task persistence
Six (Study 1)	24.59 s (17.11)	13.47 s (14.09)	19.42% (19.94)	40.06% (36.65)	16.00 (7.04)	2.85 (3.10)	39.34% (17.36)
Four (Study 2)	21.30 s (12.21)	14.46 s (12.35)	17.71% (21.17)	29.43% (35.71)	13.20 (7.12)	2.04 (2.45)	31.74% (17.06)

Note. Mean values for study time allocation (in s), restudy selections and effective restudy selections (in %), the number of concepts studied and restudied, and task persistence (in %) for both age groups. Standard deviations are in parentheses.

Table 4
Mean Study Time Allocation, Self-Score Judgments, Restudy Selections, and Performance Per Task Block

Variable	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
Sixth grade (Study 1)						
Study time allocation per concept	30.05	27.53	24.04	21.34	19.29	13.39
Self-score judgments per concept Trial 1	1.83	1.76	1.84	1.89	1.82	1.76
Self-score judgments per concept after restudy	1.91	2.16	2.27	2.19	2.05	1.79
Percentage of restudy selections	24.06	18.43	17.63	15.68	11.93	8.97
Mean performance per concept	1.62	1.60	1.72	1.67	1.49	1.28
Fourth grade (Study 2)						
Study time allocation per concept	28.60	22.23	19.88	15.15	15.90	10.30
self-score judgments per Concept Trial 1	1.54	1.48	1.60	1.51	1.52	1.45
Self-score judgments per concept after restudy	1.62	2.28	2.06	2.22	1.67	2.00
Percentage of restudy selections	21.65	13.11	15.74	11.54	11.11	8.33
Mean performance per concept	1.23	1.20	1.26	1.06	0.98	0.81

Note. Mean values for study time allocation per concept (in s), self-score judgments for trial 1 and self-score judgments after restudy per concept, the percentage of concepts restudied, and mean performance per concept for both age groups per task block.

allocation and effective restudy. Further, monitoring accuracy directly predicted performance. Of the regulation components, only task persistence predicted total task performance, and notably, of all performance predictors, task persistence was the strongest predictor.

Discussion Study 1

The first research aim was to assess relations between three components of on-task regulation. Findings are mixed (i.e., Hypothesis 1 is only partially confirmed). Although the relation between persistence and effective restudy was positive, allocation of study time was negatively related to task persistence and not related to restudy selections.

A further aim was to investigate relations between monitoring and regulation. Although monitoring judgments were related to restudy selections (confirming Hypothesis 2b), monitoring judgments were not translated into the allocation of study time during restudy (contrasting Hypothesis 2a). Further, there were no relations between self-monitoring and task persistence.

A final aim was to investigate the interrelations between monitoring accuracy, the effectiveness of regulation, and performance. Effective restudy benefited performance per studied concept, whereas persistence was the strongest predictor of total task performance. Importantly, children's monitoring accuracy directly predicted mean concept performance and total task performance. Further, monitoring accuracy predicted study time allocation and effective restudy.

In this study, participants were sixth-grade students; for a successful transition to secondary education, they need to be able to take responsibility for their learning (Dent & Koenka, 2016). Therefore, we considered it particularly relevant to investigate the interrelations between monitoring, regulation, and performance for this age group. However, during middle and late childhood, self-monitoring and self-regulation skills seem to develop (Dufresne & Kobasigawa, 1989; Metcalfe & Finn, 2013; Schneider & Löffler, 2016; Van Loon et al., 2013). Therefore, our findings may be age-sensitive. A second study was conducted with fourth-grade students to investigate potential age-related effects on the interrelations between self-monitoring, regulation, and performance. In addition to addressing the same research questions as in Study 1, findings between Study 1 and 2 are compared to address potential differences between fourth and sixth graders. We expected that the relations between monitoring, regulation, and task performance would be weaker for fourth than for sixth graders (Hypothesis 3).

Study 2

Participants

Ninety-seven fourth graders participated (M_{age} 10.37 years, *SD* 6.1 months, 53 females). Recruitment was the same as in Study 1.

Materials and Procedure

Concept selection was based on the same criteria as in Study 1. However, different concepts were used so that the difficulty

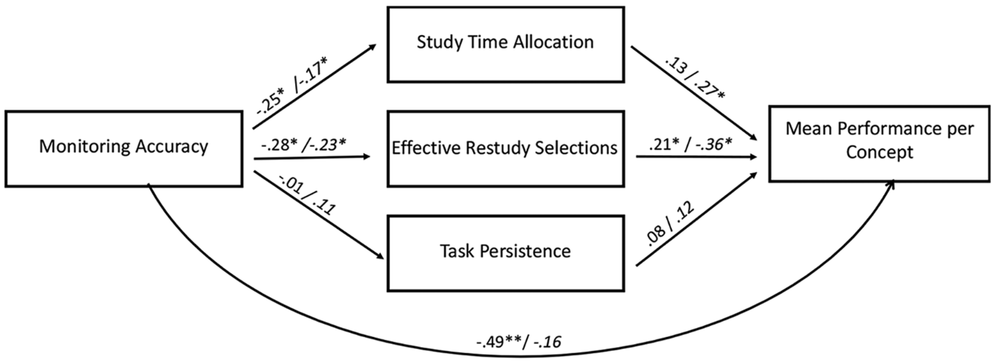
Table 5
Correlations Between Study Time Allocation, Restudy Selections, and Task Persistence

Regulation components	1	2	3	4	5	6	7
1. Study time allocation Trial 1	—	.337**	.969**	−.081	.024	−.315**	−.327**
2. Study time allocation Trial 2	.354**	—	.446**	.116	.103	−.011	.021
3. Overall study time allocation	.955**	.475**	—	.083	.160	−.342**	−.306**
4. Percentage restudy selections	.032	.011	.242*	—	.800**	−.134	.162
5. Effective restudy selections	−.032	.143	.159	.820**	—	−.007	.284*
6. Task persistence Trial 1	−.315**	.139	−.313**	−.121	−.023	—	.945**
7. Task persistence	−.315**	.138	−.257**	.126	.180	.964**	—

Note. Spearman correlations for sixth graders (Study 1) above the diagonal; Spearman correlations for fourth graders (Study 2) below the diagonal.
$^*p < .05$. $^{**}p < .01$.

Figure 2

Path Model Showing Interrelations Between Monitoring Accuracy, Regulation Components, and Mean Performance Per Concept



Note. Standardized estimates are reported for sixth graders (Study 1) and fourth graders (Study 2). Coefficients for sixth graders in normal font, coefficients for fourth graders in italics.
$^{*}p < .05$. $^{**}p < .01$.

level in Study 2 was comparable to Study 1. To ensure that children would be able to complete the digitalized concept learning task, a pilot study was conducted with ten fourth graders.

The task procedure was identical to the procedure used in Study 1. To ensure reliable coding of task performance, 31% of the test responses were scored by two independent raters. Interrater reliability was high, ICCs were 0.87.

Statistical Analyses

Analyses for Study 2 are identical to Study 1. Moreover, findings between Study 1 and Study 2 were compared to examine potential age differences in the interrelations between monitoring, regulation, and performance. To further investigate how monitoring and regulation differ between Grades 6 and 4, a logistic regression analysis was conducted with monitoring accuracy, study time allocation, effective restudy, and task persistence as predictors of grade membership. Finally, multi-group path analyses were conducted to investigate whether the interrelations between monitoring, regulation, and performance are stronger for the sixth than the fourth graders. The model fit of a constrained path model in which the path coefficients for both age groups were set to be the same was compared with an unconstrained model in which all parameters were allowed to differ between age groups. With the Chi-squared difference test, the difference between the constrained and the unconstrained model was tested. Follow-up tests then investigated for which path coefficients the 95% confidence intervals of sixth and fourth graders did not overlap, indicating statistically detectable differences between age groups.

Transparency and Openness

Ethical approval, sample size decisions, the sharing of research data and materials, and the statistical analyses for Study 2 are identical to Study 1. For the logistic regression analysis, data were analyzed using R and RStudio, version 1.4.1717.

Results

Preliminary Analyses

Table 2 shows descriptive statistics on fourth graders' test performance and monitoring judgments. As in Study 1, performance

Table 6

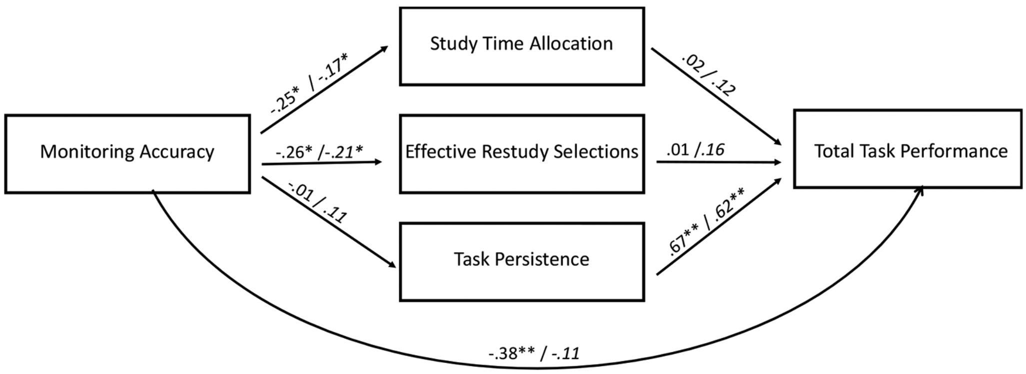
Path Coefficients and Confidence Intervals for the Path Analyses on the Interrelations Between Monitoring Accuracy, Regulation Components, and Mean Task Performance per Concept

	Sixth grade (Study 1)		Fourth grade (Study 2)	
Relations	Path coefficient (unstandardized)	95% CI	Path coefficient (unstandardized)	95% CI
Monitoring accuracy → performance	−0.74**	[−0.97 to −0.49]	−0.24	[−0.49 to 0.04]
Monitoring accuracy → study time allocation	−10.92*	[−19.29 to −3.75]	−5.02*	[−9.45 to 0.00]
Monitoring accuracy → effective restudy selections	−24.41*	[−39.30 to −7.34]	−17.47*	[−32.41 to −4.15]
Monitoring accuracy → task persistence	−0.13	[−4.27 to 4.31]	2.00	[−1.69 to 5.92]
Study time allocation → performance	0.00	[−0.00 to 0.01]	0.01*	[0.00 to 0.02]
Effective restudy selections → performance	0.00*	[0.00 to 0.01]	0.01*	[0.00 to 0.01]
Task persistence → performance	0.01	[−0.01 to 0.02]	0.01	[−0.01 to 0.03]

Note. The 95% CI values show lower and upper bounds of the bootstrap confidence intervals. Standardized coefficients are presented in Figure 2.
$^{*}p < .05$. $^{**}p < .01$.

Figure 3

Path Model Showing Interrelations Between Monitoring Accuracy, Regulation Components, and Total Task Performance



Note. Standardized estimates are reported for sixth graders (Study 1) and fourth graders (Study 2). Coefficients for sixth graders in normal font, coefficients for fourth graders in italics.
$^*p < .05$. $^{**}p < .01$.

improved through restudy, $t(96) = 12.13$, $p < .001$. Further, children were overconfident after study trial 1, calibration = 0.49 ($SD = 0.47$), $t(96) = 10.28$, $p < .001$, and after restudy, calibration = 0.42 ($SD = 0.62$), $t(64) = 5.38$, $p < .001$. There were no differences between monitoring accuracy after study trial 1 and after restudy, $t(65) = 0.54$, $p = .60$.

Table 3 shows descriptive statistics for the regulation components. Of the fourth graders, 33% did not select any items for restudy. Task persistence ranged from 8.33% to 77.08%. Table 4 shows for each task block separately an overview of monitoring judgments, regulation, and mean test performance per concept.

Interrelations Between Monitoring, Regulation, and Performance

Correlations Between Regulation Components

The Spearman correlations between the regulation components are shown in Table 5 (correlations for the fourth graders are presented below the diagonal). Although overall study time allocation per concept was weakly but positively related to the percentage of restudy selections, contrary to Hypothesis 1, study time allocation was negatively related to task persistence. There was no relation between restudy and task persistence.

Relations Between Self-Monitoring and Regulation

Contrary to Hypothesis 2a, monitoring judgments after study trial 1 were not related to subsequent allocation of restudy time, Gamma = −.20, $SD = 0.79$, $t(25) = -1.31$, $p = .20$. In line with Hypothesis 2b, monitoring judgments were related to restudy selections, Gamma = −.49, $SD = 0.67$, $t(59) = -5.58$, $p < .001$. The correlation between monitoring judgments and task persistence was low (Spearman's $\rho = .114$, $p = .27$). Further, the correlation between self-monitoring and task persistence was assessed for 86 children who were overconfident. As in Study 1, there was no relation between the degree of overconfidence and total task persistence (Spearman's $\rho = -.045$, $p > .99$).

Effects of Monitoring and Regulation on Performance

Figure 2 (coefficients for the fourth graders in italics) shows the regression coefficients for analyses investigating the effects of monitoring accuracy and regulation on mean task performance per

Table 7

Path Coefficients and Confidence Intervals for the Path Analyses on the Interrelations Between Monitoring Accuracy, Regulation Components, and Total Task Performance

	Sixth grade (Study 1)		Fourth grade (Study 2)	
Relations	Path coefficient (unstandardized)	95% CI	Path coefficient (unstandardized)	95% CI
Monitoring accuracy → performance	−4.71**	[−6.94 to −3.04]	−1.00	[−2.46 to 0.41]
Monitoring accuracy → study time allocation	−10.92*	[−18.65 to −3.18]	−5.02*	[−9.52 to −0.40]
Monitoring accuracy → effective restudy selections	−23.03*	[−37.69 to −6.38]	−16.39*	[−29.62 to −1.87]
Monitoring accuracy → task persistence	−0.13	[−4.18 to 4.04]	2.00	[−1.89 to 6.64]
Study time allocation → performance	0.01	[−0.04 to 0.04]	0.03	[−0.02 to 0.08]
Effective restudy selections → performance	0.00	[−0.02 to 0.02]	0.02	[−0.01 to 0.04]
Task persistence → performance	0.42**	[0.31 to 0.51]	0.31**	[0.21 to 0.41]

Note. The 95% CI values show lower and upper bounds of the bootstrap confidence intervals. Standardized coefficients are presented in Figure 3.
$^*p < .05$. $^{**}p < .01$.

concept. Unstandardized path coefficients and confidence intervals are presented in Table 6. There was no direct effect of monitoring accuracy on performance. Monitoring accuracy was related to study time allocation and effective restudy. Effective restudy was the strongest predictor of mean concept performance. Moreover, allocation of study time affected performance.

Figure 3 shows the standardized path coefficients for analyses addressing the effects of monitoring accuracy and regulation on total task performance; Table 7 presents the unstandardized coefficients and the confidence intervals. Monitoring accuracy affected study time allocation and effective restudy, however, there was no effect of monitoring accuracy on fourth graders' total performance. Of the regulation components, only task persistence affected total performance.

Comparing Sixth- and Fourth-Grade Students

This section compares findings from Study 1 and Study 2 to address Hypothesis 3 that interrelations between monitoring, regulation, and performance are stronger for sixth than fourth graders. We first report preliminary analyses comparing mean values for performance, monitoring, and regulation between the two age groups.

Preliminary Analyses

Performance was higher for sixth than for fourth graders after study trial 1, $t(201) = 4.41$, $p < .001$, and after restudy, $t(201) = 5.18$, $p < .001$. Monitoring judgments were higher for sixth than for fourth graders after study trial 1, $t(201) = 3.80$, $p < .001$, and after restudy, $t(141) = 2.69$, $p = .008$. The grades did not differ in the percentage of restudy selections, $t(201) = 0.59$, $p = .55$.

To identify whether and how accurate monitoring and effective regulation are different between the sixth and the fourth graders, monitoring accuracy, study time allocation, effective restudy, and task persistence were entered in a logistic regression to predict grade level (entered as a dichotomous outcome variable). The logistic regression model explained 6.5% of the variance between age groups. Table 8 shows the predictor coefficients; findings show that study time allocation and task persistence predicted grade membership. This indicates that the age groups differed in their study time allocation and task persistence. More precisely, compared to fourth graders, the sixth graders spent more time on each concept during study and got further into the task.

Table 8

Estimates of the Logistic Regression Indicating whether Monitoring Accuracy and Regulation Components Distinguish Between Fourth and Sixth Graders

Variable	Estimate	Odds ratio
Monitoring accuracy	−0.11	0.90
Study time allocation	0.02*	1.03
Effective restudy selections	0.00	1.00
Task persistence	0.06*	1.07

Note. Estimates represent the log odds of comparing Grade 6 and Grade 4 (Grade 4 is the reference group).
*$p < .05$.

Comparing Effects of Monitoring and Regulation on Performance

Multi-group analyses were computed to examine whether the path model fits the data equally well for both age groups. First, path coefficients were held equal across the two age groups. Next, path coefficients were allowed to differ across the two age groups. Then the difference between the constrained and the unconstrained model was tested. For the mean performance per concept model, constraining the path coefficients to be equal across the two age groups showed a non-significant trend ($\Delta\chi^2 = 13.27$, $\Delta df = 7$; $p = .07$), indicating that the path coefficients values tended to differ for fourth and sixth graders. Follow-up comparisons of the confidence intervals for mean performance per concept (reported in Table 6) show that the confidence intervals for the effects of monitoring accuracy on performance did not overlap. This seems to indicate that the effect of monitoring accuracy on mean performance per concept was stronger for sixth than for fourth graders.

Moreover, there was a difference between the constrained model and the non-constrained model for total task performance ($\Delta\chi^2 = 17.46$, $\Delta df = 7$; $p = .01$). This indicates that the path coefficients differ between the two age groups. Follow-up comparison of the confidence intervals (shown in Table 7) for the path coefficients for sixth graders and fourth graders shows that the confidence intervals for the effects of monitoring accuracy on performance did not overlap. This shows that the direct effect of monitoring accuracy on total task performance was stronger for sixth than fourth graders.

General Discussion

The study investigates interrelations between sixth and fourth-grade students' self-monitoring, regulation of learning, and task performance while working on a concept learning task. Multiple components of real-time regulation of learning were investigated in conjunction–marking a novel and unique approach. Children could decide how much study time they allocated, what they restudied, and how long they persisted before giving up. Most importantly, results suggest that variance in children's concept learning performance is driven by the effectiveness of regulation, particularly by task persistence. Moreover, the ability to accurately self-monitor affected performance for sixth but not fourth graders. Before further discussing the effects of monitoring and regulation on learning performance for the two age groups, we first discuss relations between components of regulation and the effects of self-monitoring on regulation.

Relations Between Regulation Components

Research using self-report scales shows that multiple regulation components are related to each other (Callan & Cleary, 2018; Pintrich et al., 1994). However, past research measuring regulation during work on tasks only investigated single regulation components; it remained unclear whether components of on-task regulation are related. Our first aim was to investigate this. Surprisingly, allocation of study time was negatively related to task persistence for both age groups. Children who spent more study time on single concepts got less far into the task. One reason for this negative relationship could be that these children may have gotten tired from spending much cognitive effort on individual items (Chevalier, 2018). This may have caused them to stop at an earlier point in the task than children who more quickly moved on to the following concepts. Further, the overall number of restudy selections was not related to the allocation of study time and task persistence. These findings imply that, when measuring real-time, on-task regulation of learning, regulation has a more multidimensional nature than when using self-report measures.

Interestingly, for sixth but not for fourth graders, there was a relation between task persistence and effective restudy. Sixth-grade children who got further into the task also more strategically selected the not-yet-learned items for restudy. This may indicate that children start using multiple self-regulation strategies simultaneously to proceed toward learning goals between 10 and 12 years of age. Future research could further investigate (a) whether relations between regulation components depend on the measurement approach (i.e., self-report versus on-task measures) and (b) to what extent relations between multiple components of regulation increase with age.

Interrelations Between Self-Monitoring, Regulation, and Performance

As hypothesized, self-monitoring judgments were strongly related to restudy selections (confirming Van Loon et al., 2017); more accurate monitoring led to more effective restudy. Moreover, the children who more accurately monitored their performance (i.e., children who were less overconfident) also allocated more study time per concept. However, when investigating the intra-individual relations between monitoring judgments and study time allocation on the item level, in contrast to previous findings (e.g., Destan et al., 2014), children did not necessarily allocate more study time to items they judged as not-yet-well-learned. Possibly, in the present study, there were individual differences between children in their strategic approaches to study time allocation. One approach could be to allocate most of the restudy time to items that were identified as least learned (i.e., received low monitoring judgments, Thiede & Dunlosky, 1999). In the study by Destan et al. (2014), children mainly seem to have used this approach. However, when task items are perceived as difficult to learn and too challenging to master, participants have been shown to move quickly to easier items (Metcalfe & Kornell, 2005). Then, learning performance may benefit most when study time is strategically allocated to the items that are judged as more learnable. Approaches to study time allocation seem to depend on individual differences in participants' moment-to-moment learning experiences and learning goals (Ariel et al., 2009; De Bruin et al., 2020). Future research could assess learning goals before and during work on a task and perceptions about item difficulty and ease-of-learning to further understand variance in study time allocation.

Although relations between monitoring accuracy and task persistence have not been explicitly addressed in previous research, it has been suggested that overconfidence may be beneficial for children's motivation to persist (Bjorklund & Beers, 2016; Lipko et al., 2009; Shin et al., 2007). However, the present findings imply that overestimating performance was not beneficial for task persistence (i.e., for both age groups, there were no relations between monitoring accuracy and task persistence). Moreover, for the oldest age group, overconfidence was disadvantageous for concept learning performance. This contrasts with Shin et al. (2007), who found that children who overestimated their memory had better recall performance than children who more accurately monitored their memory. However, Shin et al. (2007) also showed that for older children (age 9), positive relations between overconfidence and performance were less pronounced than for 6–7-year-olds. The present research findings suggest that during late childhood, there may be a turning point in the effects of monitoring accuracy on performance. It seems that around the age of 10, children's monitoring does not directly affect performance, even though accurate monitoring leads to more allocation of study time and more effective restudy selections. That is, for the younger age group, being overconfident did not harm task performance (although it has to be noted that there were no benefits of overconfidence). However, during late childhood, around the age of 12, monitoring accuracy becomes a strong predictor of performance. For the older age group, overestimating performance was clearly disadvantageous for learning.

Conclusions about relations between regulation and performance may be affected by the measure of performance, that is, whether performance is assessed per studied concept or for the entire task. Task persistence was a strong predictor of total task performance. For mean concept performance, particularly effective restudy had benefits for both age groups. Moreover, for fourth but not for sixth graders, study time allocation affected mean performance per concept. This may indicate that for younger children, putting in extra time may be useful to improve concept learning. In contrast, for older children, particularly accurate monitoring seems to be important.

The theoretical metacognition model by Nelson and Narens (1990) suggests directional relations between self-monitoring, regulation, and performance. The present research confirms these theoretical ideas for a sample of participants in middle and late childhood. Moreover, the findings support the assumptions of the COPES model (Winne & Hadwin, 1998) for school children. Self-regulation of learning involves self-monitoring (i.e., evaluating) the discrepancy between a current state of learning (i.e., the product) and the externally set goal to learn the concepts as well as possible (i.e., the standards). Based on self-monitoring, children then take actions (i.e., operate); our findings show that accurate monitoring positively affects study time allocation and effectively restudying of not-yet-learned concepts. Moreover, a person's age (in the COPES model, age could be considered a condition) seems to affect the relations between monitoring accuracy and performance.

Developmental Differences in Interrelations Between Monitoring, Regulation, and Performance

A comparison between age groups in monitoring and regulation showed that, in comparison to fourth graders, sixth graders (a) more strongly based restudy selections on their monitoring judgments; (b) put in more study time per concept; and (c) had a higher level of task persistence. This indicates that multiple skills to regulate learning seem to develop between 10 and 12 years. Previous research mainly focused on the age-related improvement of children's skills to translate monitoring judgments into restudy, referred to as monitoring-based-control (Dufresne & Kobasigawa, 1989; Metcalfe & Finn, 2013; Schneider & Löffler, 2016; Van Loon et al., 2013). Our findings confirm the development of monitoring-based control. Moreover, the present study shows that study time allocation skills and task persistence develop during late childhood. Findings that the percentage of restudy selections did not differ between fourth and sixth graders indicate that the younger children were equally motivated to study the items as the older age group. However, the younger children may not have had the ability to continue as long; they may have gotten tired, which may have caused them not to allocate the same amount of study time per concept as sixth graders and stop the task at an earlier time point.

Although task persistence was the strongest predictor of total performance in both age groups, for sixth graders, monitoring accuracy was a further predictor of performance (with a strong effect size), and monitoring directly and independently predicted performance even after accounting for the effects of regulation. For fourth graders, monitoring did not have direct effects on performance. This implies that accurate self-monitoring becomes crucial for learning during late childhood.

Limitations, Future Directions, and Conclusions

A novel contribution of this research is that children's strategic regulation of learning is investigated with multiple on-task measures. However, findings need to be interpreted with care. The present task was a language learning task, for which concept definitions were the primary targets of learning. The methodology was based on previous research investigating monitoring and regulation when learning concepts (e.g., Rawson & Dunlosky, 2007; Lipko et al., 2009; Van Loon et al., 2017). A specific advantage of this task is that the accuracy of monitoring and the effectiveness of regulation can be assessed per item. For a wide variety of educational subjects, students are required to learn and apply abstract concepts (Rawson et al., 2015), which makes findings relevant for learning in the educational context. However, a limitation of the concept learning test is that it assessed to what extent students remembered the previously learned definition, rather than investigating higher-order skills related to the application of learned content. As findings may be task- and test-specific, the results warrant replication with different learning tasks and higher-level measures of performance.

The theoretical COPES model (Winne & Hadwin, 1998) emphasizes that internal and external conditions affect monitoring, regulation, and performance. In the present study, we addressed how participants' age (which could be considered a condition) affects the relations between monitoring, regulation, and performance. In the COPES model, task characteristics, characteristics of the individual (e.g., motivation and affect), and the social context are all considered conditions. Addressing the effects of different conditions would have gone beyond the research scope to investigate relations between monitoring, regulation, and performance. However, future research could include variables related to cognitive capacity (e.g., working memory), motivation (e.g., self-efficacy, goal-setting), and effort (e.g., cognitive load) to better understand how these factors affect monitoring, regulation, and performance.

Interestingly, total task performance did not seem to benefit from allocating study time to individual concepts. However, in the present study, all concepts were unrelated, so quickly skipping a concept did not affect learning a new concept. For higher-order learning tasks (e.g., expository text comprehension or problem-solving), specific information needs to be understood before learning subsequent information. Then, it may be rather useful to allocate study time until the information is fully understood before moving to the next subtasks. Future research could address how the effects of study time allocation and task persistence may depend on the type of task.

Moreover, task persistence appeared to be a strong predictor of total task performance, and effective restudy predicted mean performance per learned concept. Participants were informed that the task goal was to learn as many concept definitions as thoroughly as possible. Although participants rather decided to restudy items for which monitoring judgments were low than items for which they gave high judgments, they still often chose not to restudy items they judged as not well learned. Further, there were large individual variances in persistence such that some participants gave up at an early point, whereas other participants were sticking much longer to the task. Participants may have had different reasons for their decisions about whether or not to allocate study time, restudy concepts, and persist with the task. A limitation of this research is that we did not obtain insights into the underlying reasons for these individual differences in regulation. Most likely students had different initial goals for how long they were willing to study the concepts. There was no external reward (such as getting a higher grade) for restudying and persisting, and possibly, participants who restudied effectively and persisted were intrinsically motivated to continue. For instance, they may have found the task more interesting than the persons who stopped at an earlier point. Further, participants who learned more concepts and restudied these more often may have had the belief that investing effort would benefit performance.

The research task was designed so that it matched regulation of learning in the actual educational context, where students are presented with external goals and use their internal goals to make choices about how to work on tasks. The present findings indicate that the communicated goal "to learn as many concept definitions as well as possible" was adopted, at least to some extent. The concepts judged as less-learned were more often restudied than concepts judged as well-learned. This indicates that learners tried to improve their performance through restudying. Further, findings on improved performance after restudy also show that participants strategically used restudy to improve concept learning. Notably, monitoring accuracy and the components of regulation explained an extensive part of the variance in performance. However, it is important to note that individual differences between participants in terms of personal goals (despite giving them a pre-set task goal) are likely to influence monitoring, regulation, and performance. Future research could investigate how internal conditions (e.g., individual differences in personal goals, motivation, and beliefs) and external conditions (e.g., task rewards or social pressure) influence on-task regulation.

In sum, this research brings evidence that children's on-task regulation of learning is a multidimensional construct. From fourth grade onwards, effective regulation is essential for learning; effective restudy selections predict mean performance per concept, and task persistence seems to be a strong predictor of total task performance. Age-related improvements were shown for monitoring-based restudy selections, study time allocation, and task persistence. Interrelations between monitoring, regulation, and performance were most pronounced for the sixth graders, and importantly, monitoring accuracy was a strong and independent predictor of performance for sixth but not for fourth graders. This seems to indicate that for fourth graders (around 10 years of age), it may be too early to emphasize monitoring accuracy in the educational context. For this age group, it may make more sense to address how actual regulation can be effective while working on tasks. However, during late childhood (when children are around 12 years of age), addressing and training self-monitoring should have a prominent role in the school curricula.

References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology*, *17*(1), 18–32. https://doi.org/10.1037/a0022086

Andersson, H., & Bergman, L. R. (2011). The role of task persistence in young adolescence for successful educational and occupational attainment in middle adulthood. *Developmental Psychology*, *47*(4), 950–960. https://doi.org/10.1037/a0023786

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*(3), 432–447. https://doi.org/10.1037/a0015928

Bjorklund, D. F., & Beers, C. (2016). The adaptive value of cognitive immaturity: Applications of evolutionary developmental psychology to early education. In D. C. Geary & D. B. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp. 3–32). Springer. https://doi.org/10.1007/978–3-319-29986-0_1

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, *54*(2), 199–231. https://doi.org/10.1111/j.1464-0597.2005.00205.x

Callan, G. L., & Cleary, T. J. (2018). Multidimensional assessment of self-regulated learning with middle school math students. *School Psychology Quarterly*, *33*(1), 103–111. https://doi.org/10.1037/spq0000198

Chevalier, N. (2018). Willing to think hard? The subjective value of cognitive effort in children. *Child Development*, *89*(4), 1283–1295. https://doi.org/10.1111/cdev.12805

Cleary, T. J., & Zimmerman, B. J. (2004). Self-regulation empowerment program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools*, *41*(5), 537–550. https://doi.org/10.1002/pits.10177

Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning*, *1*(3), 229–247. https://doi.org/10.1007/s11409-006-9002-5

De Bruin, A. B. H., Roelle, J., Carpenter, S. K., & Baars, M. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educational Psychology Review*, *32*(4), 903–915. https://doi.org/10.1007/s10648-020-09576-4

De Bruin, A. B. H., & Van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom [editorial material]. *Learning and Instruction*, *22*(4), 245–252. https://doi.org/10.1016/j.learninstruc.2012.01.003

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review*, *28*(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

Destan, N., Hembacher, E., Ghetti, S., & Roebers, C. M. (2014). Early metacognitive abilities: The interplay of monitoring and control processes in 5-to 7-year-old children. *Journal of Experimental Child Psychology*, *126*, 213–228. https://doi.org/10.1016/j.jecp.2014.04.001

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*(3), 231–264. https://doi.org/10.1007/s11409-008-9029-x

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology*, *47*(2), 274–296. https://doi.org/10.1016/0022-0965(89)90033-7

Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: Problems and pitfalls. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 23–37). Oxford University Press.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, *24*, 58–61. https://doi.org/10.1016/j.learninstruc.2012.05.002

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, *24*(2), 462–475. https://doi.org/10.1037/a0014417

Howard, M. C., & Crayne, M. P. (2019). Persistence: Defining the multidimensional construct and creating a measure. *Personality and Individual Differences*, *139*, 77–89. https://doi.org/10.1016/j.paid.2018.11.005

Ladd, G. W., & Dinella, L. M. (2009). Continuity and change in early school engagement: Predictive of children's Achievement trajectories from first to eighth grade? *Journal of Educational Psychology*, *101*(1), 190–206. https://doi.org/10.1037/a0013153

Lawson, M. A., & Lawson, H. A. (2013). New conceptual frameworks for student engagement research, policy, and practice. *Review of Educational Research*, *83*(3), 432–479. https://doi.org/10.3102/0034654313480891

Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology*, *103*(2), 152–166. https://doi.org/10.1016/j.jecp.2008.10.002

Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning*, *8*(1), 19–46. https://doi.org/10.1007/s11409-013-9094-7

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*(4), 463–477. https://doi.org/10.1016/j.jml.2004.12.001

Moilanen, K. L. (2007). The adolescent self-regulatory inventory: The development and validation of a questionnaire of short-term and long-term self-regulation. *Journal of Youth and Adolescence*, *36*(6), 835–848. https://doi.org/10.1007/s10964-006-9107-9

Mokrova, I. L., O'Brien, M., Calkins, S. D., Leerkes, E. M., & Marcovitch, S. (2013). The role of persistence at preschool age in academic skills at kindergarten. *European Journal of Psychology of Education*, *28*(4), 1495–1503. https://doi.org/10.1007/s10212-013-0177-2

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, *26*, 125–173. https://doi.org/10.1016/S0079-7421(08)60053-5

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (Vol. 13, pp. 1–25). MIT Press.

Oeri, N., & Roebers, C. M. (2021). Task persistence in kindergarten children: Disentangling age from schooling effects. *British Journal of Developmental Psychology*, *39*(1), 217–230. https://doi.org/10.1111/bjdp.12358

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Into Practice*, *41*(4), 219–225. https://doi.org/10.1207/s15430421tip4104_3

Pintrich, P. R., Roeser, R. W., & De Groot, E. A. (1994). Classroom and individual differences in early adolescents' motivation and self-regulated learning. *The Journal of Early Adolescence*, *14*(2), 139–161. https://doi.org/10.1177/027243169401400204

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*(4–5), 559–579. https://doi.org/10.1080/09541440701326022

Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review*, *27*, 483–504. https://doi.org/10.1007/s10648-014-9273-3

Rinne, L. F., & Mazzocco, M. M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoS One*, *9*(7), Article e98663. https://doi.org/10.1371/journal.pone.0098663

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schneider, W., & Löffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dunlosky & R. A. Bjork (Eds.), *The Oxford handbook of metamemory* (pp. 191–517). Oxford University Press. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Shin, H., Bjorklund, D. F., & Beck, E. F. (2007). The adaptive nature of children's Overestimation in a strategic memory task. *Cognitive Development*, *22*(2), 197–212. https://doi.org/10.1016/j.cogdev.2006.10.001

Son, L. K., & Kornell, N. (2008). Research on the allocation of study time: Key studies from 1890 to the present (and beyond). In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 333–351). Taylor & Francis Group.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66–73. https://doi.org/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*(4), 1024–1037. https://doi.org/10.1037/0278-7393.25.4.1024

Thiede, K. W., Redford, J. S., Wiley, J., & Griffin, T. D. (2017). How restudy decisions affect overall comprehension for seventh-grade students. *British Journal of Educational Psychology*, *87*(4), 590–605. https://doi.org/10.1111/bjep.12166

Van Loon, M., De Bruin, A., Leppink, J., & Roebers, C. (2017). Why are children overconfident? Developmental differences in the implementation of accessibility cues when judging concept learning. *Journal of Experimental Child Psychology*, *158*, 77–94. https://doi.org/10.1016/j.jecp.2017.01.008

Van Loon, M. H., De Bruin, A. B. H., Van Gog, T., & Van Merriënboer, J. J. G. (2013). The effect of delayed-JOLs and sentence generation on children's Monitoring accuracy and regulation of idiom study. *Metacognition and Learning*, *8*(2), 173–191. https://doi.org/10.1007/s11409-013-9100-0

Van Loon, M. H., & Roebers, C. M. (2017). Effects of feedback on self-evaluations and self-regulation in elementary school. *Applied Cognitive Psychology*, *31*(5), 508–519. https://doi.org/10.1002/acp.3347

Van Loon, M., & Van de Pol, J. (2019). Judging own and peer performance when using feedback in elementary school. *Learning and Individual Differences*, *74*, Article 101754. https://doi.org/10.1016/j.lindif.2019.101754

Van Loon, M., & Oeri, N. S. (2022, October 25). *Examining on-task regulation in school children: Interrelations between monitoring, regulation, and task performance*. https://osf.io/y25zk/

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Erlbaum.

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Erlbaum.

Zamary, A., & Rawson, K. A. (2018). Are provided examples or faded examples more effective for declarative concept learning? *Educational Psychology Review*, *30*, 1167–1197. https://doi.org/10.1007/s10648-018-9433-y

Zhou, Q., Hofer, C., Eisenberg, N., Reiser, M., Spinrad, T. L., & Fabes, R. A. (2007). The developmental trajectories of attention focusing, attentional and behavioral persistence, and externalizing problems during school-age years. *Developmental Psychology*, *43*(2), 369–385. https://doi.org/10.1037/0012-1649.43.2.369

Received October 1, 2021
Revision received September 23, 2022
Accepted September 27, 2022 ■