

# Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration

Citation for published version (APA):

van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. J. G. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning and Instruction*, 24, 15-25. <https://doi.org/10.1016/j.learninstruc.2012.08.005>

## Document status and date:

Published: 01/04/2013

## DOI:

[10.1016/j.learninstruc.2012.08.005](https://doi.org/10.1016/j.learninstruc.2012.08.005)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

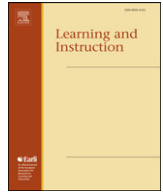
[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



## Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration

Mariëtte H. van Loon<sup>a,\*</sup>, Anique B.H. de Bruin<sup>a,b</sup>, Tamara van Gog<sup>b</sup>, Jeroen J.G. van Merriënboer<sup>a</sup>

<sup>a</sup> Maastricht University, Department of Educational Development & Research, Graduate School of Health Professions Education, P.O. Box 616, 6200 MD Maastricht, The Netherlands

<sup>b</sup> Erasmus University Rotterdam, Department of Psychology, P.O. Box 1783, 3000 DR Rotterdam, The Netherlands

### ARTICLE INFO

#### Article history:

Received 29 January 2012

Received in revised form

18 August 2012

Accepted 21 August 2012

#### Keywords:

Calibration

Judgment accuracy

Regulation of study

Children

Overconfidence

### ABSTRACT

The study investigated whether activation of inaccurate prior knowledge before study contributes to primary-school children's commission errors and overconfidence in these errors when learning new concepts. Findings indicate that inaccurate prior knowledge affects children's learning and calibration. The level of children's judgments of learning for recall responses for which they would not receive credit was inappropriately high after activation of inaccurate prior knowledge.

Moreover, results showed that activation of inaccurate prior knowledge was not only detrimental for monitoring judgments during learning, but also for calibration accuracy after test taking. When judging the quality of their recall responses on the posttest, children were more overconfident when they had activated inaccurate prior knowledge. Also, the children often discarded concepts from further study after activation of inaccurate prior knowledge. These results suggest that in order to improve self-regulated learning, it may be important to detect inaccuracies in children's prior knowledge.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

Learning and understanding new concepts presented in a text is an activity that young learners have to perform frequently (Lipko, Dunlosky, Hartwig, et al., 2009). In order to study efficiently, students have to accurately monitor the quality of their learning while studying, that is, by correctly judging the extent to which the studied information has been understood (e.g. Hadwin & Webster, 2013; Nelson & Narens, 1990; Thiede, Anderson, & Theriault, 2003). This affects the efficiency of their learning, because without accurate monitoring, learners might not spend enough time studying concepts they think they know, but actually have not yet learned. However, monitoring the quality of learning is not only relevant during the study phase. It is also important for learners to accurately evaluate the quality of their test performance, so that when taking in-between self-tests, they can use this information to regulate further study.

Unfortunately, research on monitoring judgments made during learning (i.e., Judgments of Learning – JOLs) and on monitoring test performance (i.e., Self-Score Judgments – SSJs) has shown that both tend to be inaccurate (Dunlosky, Rawson, & Middleton, 2005; Nelson & Dunlosky, 1991). The present study aimed to investigate whether poor monitoring accuracy might in part be caused by the

accessibility of inaccurate prior knowledge. We will first give an overview of research on the accuracy of JOLs and SSJs, before examining why activating inaccurate knowledge prior to study might contribute to inaccurate monitoring.

The research on JOL accuracy has mainly focused on relative accuracy, which is usually measured by means of intra-individual correlations between JOLs and performance on test items (Maki, Shields, Wheeler, & Zacchilli, 2005; Nelson, 1984; Rhodes & Tauber, 2011; Schraw, Kuch, & Gutierrez, 2013). When providing JOLs about studied concepts, the higher the JOLs are, the more confident the learner is to retrieve the correct meaning of the concept on the test. The correlation between JOLs and performance, with values ranging between  $-1$  and  $+1$ , indicates whether a person can discriminate between items that are more or less likely to be recalled on a test. A highly positive correlation between JOLs and performance shows accurate monitoring, whereas a value of zero indicates that there is no association between predictions and performance. Research has shown that when JOLs are made immediately after studying an item, correlations often do not exceed .25, indicating that students have problems discriminating between items which they have or have not understood (Dunlosky & Lipko, 2007). When regulating further study, learners typically select items for restudy for which they gave low JOLs (Metcalf & Finn, 2008; Nelson, Dunlosky, Graf, & Narens, 1994). Thus, when JOLs are inaccurate, students cannot accurately decide which items should receive the highest priority for additional study.

\* Corresponding author. Tel.: +31 43 3885720; fax: +31 43 3885779.

E-mail address: [m.vanloon@maastrichtuniversity.nl](mailto:m.vanloon@maastrichtuniversity.nl) (M.H. van Loon).

Accuracy of Self-Score Judgments (SSJs) is usually measured in terms of calibration, or absolute accuracy, by comparing the correspondence between an SSJ for an item and performance on that item. SSJs require learners to provide an estimate regarding the quality of their response (i.e., their recall, Dunlosky et al., 2005). When providing SSJs for the recall of concepts, learners usually indicate on a scale whether they expect to receive no, partial, or full credit for their response (e.g., Dunlosky et al., 2005; Rawson & Dunlosky, 2007).

By calculating the difference between the SSJs and performance, the measure of SSJ calibration accuracy provides insight into the match between a learner's perception of performance and the actual level of performance. When comparing SSJs with actual test scores, researchers acquire insight into the level of over- or underconfidence of a learner. The smaller the difference between SSJs and actual scores, the better the calibration of the learner is (Hacker, Bol, & Keener, 2008). A person is overconfident if the calculated discrepancy between SSJs and test scores is a positive value, and underconfident if this value is negative. For example, if a learner judges the quality of a recall response as fully correct (1 credit) and the actual test score for that item is partially correct (.5 credit), this shows that the student was overconfident. On the other hand, if the learner would have provided an SSJ of 0 credit and the actual test response would receive .5 credit, this person would be underconfident.

Unfortunately, research has shown that learners are often inaccurate when making SSJs. Both adults and younger learners tend to be overconfident (Flavell, 1970; Koriat, 2011) when monitoring the quality of recall responses that are entirely incorrect, referred to as commission errors. Even though these responses are not awarded credit, learners often expect to receive partial or full credit for their commission errors (Dunlosky et al., 2005; Lipko, Dunlosky, & Merriman, 2009; Rawson & Dunlosky, 2007). Findings with eighth-grade learners showed that as much as 43% of their recall responses consisted of commission errors (Lipko, Dunlosky, Hartwig, et al., 2009). When regulating further study, this tends to be inefficient as learners usually discard items for which they believe their recall response was correct. Commission errors thus contribute to inefficient study regulation to a large extent (Dunlosky, Hartwig, Rawson, & Lipko, 2011; Dunlosky & Rawson, 2012).

The cue utilization framework (Koriat, 1997) provides an explanation for learner's poor accuracy, and overconfidence when monitoring learning and performance. Presumably, poor accuracy results when learners base their judgments on cues that are not valid indicators of test performance. According to the accessibility theory (Koriat, 1993), learners often derive their monitoring judgments from the amount of accessible information that comes to mind. Koriat (1995) found that adult learners' confidence increased when the accessibility of information increased (i.e. the presence of information in memory), and this increase in confidence was independent of the actual correctness of their test response (i.e. the quality). Accessibility might not be a valid cue when monitoring learning, because learners would focus on the quantity of information in memory, instead of the quality, which is what is being judged by teachers when students take a test (Dunlosky et al., 2005; Koriat, 1993, 1995; Koriat & Levy-Sadot, 2001; Rawson & Dunlosky, 2007).

Studies have shown that monitoring accuracy can improve when learners' attention is focused on valid cues. Improvements in the relative accuracy of JOLs have been found when they are provided at a short delay after learning, rather than immediately during learning (Koriat & Shitzer-Reichert, 2002; Nelson & Dunlosky, 1991; Rhodes & Tauber, 2011; Schneider, Visé, Lockl, & Nelson, 2000). However, the measure of relative JOL accuracy shows only whether participants can discriminate between items

they studied, it does not provide insight into the *level* of learners' overconfidence in their commission errors (Schraw, 2009). A measure of calibration, or absolute accuracy, would be needed to determine to what extent learners are able to accurately monitor the quality of their recall response on each item.

The few studies that have investigated calibration accuracy of young learners' SSJs have shown that children show a high level of overconfidence (Barnett & Hixon, 1997; Huff & Nietfeld, 2009; Lipko, Dunlosky, Hartwig, et al., 2009). An important factor that might affect the calibration accuracy of commission errors, but which has, to our knowledge, not yet received any attention, is accessibility of prior knowledge. We know that learners' prior knowledge has a large influence on what they understand and remember when learning from text (Alexander & Jetton, 2000). Several studies have shown that activating prior knowledge before studying has positive effects on learning (e.g. Ozuru, Dempsey, & McNamara, 2009; Wetzels, Kester, Van Merriënboer, & Broers, 2011). However, inaccurate prior knowledge may have a detrimental effect not only on learning but also on monitoring, for instance, when learners have inaccurate ideas about the meaning of certain concepts to be studied.

Research has found that when the inaccuracy of prior knowledge is not directly identified and addressed, it can be resistant to change and hinder the learning of new information (Dochy, Segers, & Buehl, 1999; Lipson, 1982; Van den Broek & Kendeou, 2008). For instance, Lipson (1982) showed that when children activated inaccurate prior knowledge before studying a text that contradicted their prior knowledge, they still used their inaccurate prior knowledge when answering questions after study. It has been suggested that learning outcomes might actually be better when learners are not able to activate prior knowledge than when they activate inaccurate prior knowledge (Lipson, 1982).

In a study examining the effect of prior knowledge on calibration, it has been found that learners reported using their prior knowledge when monitoring learning (Dinsmore & Parkinson, 2013). However, learners' prior knowledge may either be absent, completely inaccurate, partially accurate, or fully accurate. Inaccuracy of prior knowledge may be especially problematic when learners use their prior knowledge in monitoring. Activation of inaccurate prior knowledge might lead to accessibility of incorrect information both during study and during test taking. If learners are unaware of the inaccuracy of their prior knowledge, they might judge these responses as partially or fully correct, leading to overconfidence. In sum, this study intends to investigate whether inaccurately activated prior knowledge might lead to overconfidence in commission errors, more than when learners do not have any prior knowledge accessible. Note that, of course, it is also possible that a learners' prior knowledge is partially correct or fully correct. However, in the present study we focus on unsuccessful prior knowledge activation, because to improve learning, insight into monitoring and regulation processes for incorrect responses is far more important than insight into metacognitive processes for correct responses (Krebs & Roebbers, 2010).

In the present study, primary-school learners studied a list of concepts with their meanings. Prior to this, they were required to activate any prior knowledge they had about these concepts. After studying the meaning of the concepts, the students were asked to monitor their learning of these concepts (by providing JOLs and SSJs), and to regulate further study (by selecting concepts for restudy).

In line with the findings of Lipson (1982), first of all we expected that activation of inaccurate prior knowledge before study would persist on the recall test, thus leading to more commission errors at the recall test compared to items for which no prior knowledge was available (Hypothesis 1).

Second, activation of inaccurate prior knowledge before study was expected to affect monitoring accuracy. JOLs were expected to indicate more overconfidence after activation of inaccurate prior knowledge compared to no activation of prior knowledge (Hypothesis 2a). Moreover, consistent with the findings of Lipko, Dunlosky, Hartwig, et al. (2009) and Lipko, Dunlosky, and Merriman (2009), we expected that SSJs would show overconfidence. More specifically, if activation of inaccurate prior knowledge is detrimental for calibration, we would expect that SSJs would show a higher level of overconfidence after activation of inaccurate prior knowledge at the pretest than when children were not able to activate any prior knowledge (Hypothesis 2b).

Third, if accessibility of inaccurate prior knowledge has a negative effect on monitoring accuracy, then it is also likely that regulation of study is negatively affected. When investigating regulation regarding concepts for which learners would not show correct understanding at the test, we would expect less efficient regulation after activation of inaccurate prior knowledge. That is, learners are expected to prematurely discard these concepts from further study more often after activation of inaccurate prior knowledge than after not being able to activate any prior knowledge (Hypothesis 3).

There is evidence that monitoring accuracy and overconfidence might be affected by developmental factors (Schneider & Pressley, 1997). Therefore, we conducted two studies to investigate developmental differences, by using similar procedures in the two studies but involving children of different age cohorts. In Study 1, participants were in third- and fourth-grade (8/9-year-old), whereas Study 2 involved fifth- and sixth-grade (10/11-year-old) children. At a young age, relative accuracy of JOLs can already be high, because children seem to be able to accurately discriminate between items (e.g. Dufresne & Kobasigawa, 1989; Koriatic & Shitzer-Reichert, 2002; Schneider et al., 2000). However, research on calibration has shown that 8/9-year-old children tend to be more overconfident and often overestimate their memory performance than 10/11-year-old children (e.g. Flavell, 1970; Krebs & Roebbers, 2010, 2011; Schneider & Pressley, 1997). Krebs and Roebbers (2010) showed that 10/11-year-old children can monitor more accurately, and show better calibration for questions on which they cannot provide a correct answer than 8/9-year-old children. Moreover, research indicates that older primary-school learners show better regulation abilities than younger learners (e.g. De Bruin, Thiede, Camp, & Redford, 2011; Krebs & Roebbers, 2010). So we addressed our hypotheses, keeping the nature of the learning task similar across the two studies, but the concepts differed to ensure that the materials were age appropriate.

## 2. Study 1

### 2.1. Method

#### 2.1.1. Participants

One-hundred-and-three primary-school children from two third-grade ( $n = 51$ ) and two fourth-grade classrooms in The Netherlands ( $n = 52$ ) participated in this study (51.5% male, 48.5% female;  $M = 8.6$  years,  $SD = .8$ ). Most of the children came from middle-class families and all participants exhibited sufficient Dutch language abilities to follow school lessons and to understand our instructions.

#### 2.1.2. Materials

The concept study task was a pen-and-paper task. Based on a study on lexical richness of young learners in the Dutch language (Vermeer, 2000), and a pilot study with four learners in the same age group as the participants, Dutch concepts were selected. These concepts were assumed to be unknown for learners in the age

cohort under investigation, thus learners were assumed to have very little accurate prior knowledge about the items. Fig. 1 shows a representation of the materials and the procedure of the study.

The pretest consisted of a list of the to be studied concepts on one page, containing a space to write down their response about the meaning next to each concept.

The concept study task consisted of 20 concepts which were presented with their meaning and an example sentence in which the concept was used. The meanings and example sentences were taken from a dictionary for young learners (Verburg, 2008). JOLs were provided by marking a square on an 11-point thermometer adapted from Koriatic and Shitzer-Reichert (2002), ranging from “certain I will not recall the meaning” (0) to “certain I will recall the meaning” (10).

When the participants regulated their study, they selected items for restudy on a page that displayed the concepts in a  $10 \times 2$  array. The JOLs were not visible when participants selected concepts for restudy.

The posttest (recall test) was the same as the pretest. Self-score judgments (SSJs) were provided by marking on a square which was depicted next to the tested concept whether the participants expected their recall response to be incorrect, partially correct or fully correct (in line with e.g., Lipko, Dunlosky, Hartwig, et al., 2009).

All subtasks were presented in a booklet, with blank pages in between the subtasks. The order of the concepts was randomized across children and across subtasks (i.e., pretest, study task, JOL task, restudy selection, recall test, and SSJ task). The obtained results were not likely to be influenced by order of item presentation because children received different versions of the concept tasks, and the order of concepts was different for every subtask.

#### 2.1.3. Procedure

Children were tested in the classroom, the duration of the session was approximately 1 h. The task began with the experimenter informing the participants that they would receive a booklet with subtasks, and that blank pages indicated the end of subtasks. Children were asked to wait when they came to a blank page until the experimenter announced they could turn that page. The experimenter then explained that they would study concepts, and provided them with an example of a concept and an example of the recall test. Participants were informed that the aim of studying was to show their understanding of the concepts during testing. After this explanation, children received the booklet and were asked to provide their names on the first page.

The concept task consisted of six consecutive phases which are depicted in Fig. 1. The first phase was a pretest of prior knowledge. Participants saw the 20 concepts listed, and were asked to write down the meaning of the concept if they knew it, and to leave the space blank if they did not know its meaning.

After 10 min they came to the second phase and were asked to study the materials. At the start of this study phase, the experimenter showed that the task consisted of 20 concepts, and told learners they would have 10 min to study these concepts. After the study phase, children performed an unrelated filler task (i.e., finding the differences between two pictures that were almost identical but differed on 6 small details) to insert a delay between study and providing JOLs.

In phase 3, children were asked to provide JOLs. First, they were shown the example concept again, and received instructions about the JOL scale. Use of the thermometer was explained and practiced. The experimenter discussed with the children what the various points on the thermometer meant in terms of likelihood to recall the meaning of a concept on the test. Then, participants were asked to provide JOLs for each studied concept in their booklet by marking the corresponding square of choice on the 11-point thermometer.


<b>1. Pretest</b>	<p>What does <b>fossil</b> mean?</p> <p>What does <b>cider</b> mean?</p>						
<b>2. Study</b>	<p>What does <b>fossil</b> mean? - Fossils are the remains of animals or plants from the past.</p> <p><i>Sometimes people find a fossil when they dig deep under the ground.</i></p>						
<b>3. Judgments of Learning</b>	<p>How confident are you that you will recall the meaning of:</p> <p style="text-align: center;"><b>Fossil</b></p> <p style="text-align: center;">not at all <span style="float: right;">sure</span></p> 						
<b>4. Selection for Restudy</b>	<p>Which words would you like to study again?</p> <table style="width: 100%; text-align: center;"> <tr> <td><b>Cider</b></td> <td><b>Lecture</b></td> </tr> <tr> <td><b>Pioneer</b></td> <td><b>Fossil</b></td> </tr> </table>	<b>Cider</b>	<b>Lecture</b>	<b>Pioneer</b>	<b>Fossil</b>		
<b>Cider</b>	<b>Lecture</b>						
<b>Pioneer</b>	<b>Fossil</b>						
<b>5. Test</b>	<p>What does <b>fossil</b> mean?</p> <p>.....</p> <p>.....</p>						
<b>6. Self-Score Judgments</b>	<table style="width: 100%; text-align: center;"> <tr> <td><b>No Credit</b></td> <td><b>Partial Credit</b></td> <td><b>Full Credit</b></td> </tr> <tr> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> <td><input type="checkbox"/></td> </tr> </table>	<b>No Credit</b>	<b>Partial Credit</b>	<b>Full Credit</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>No Credit</b>	<b>Partial Credit</b>	<b>Full Credit</b>					
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					

Fig. 1. Example of concept task items (original in Dutch).

The fourth phase began after the children provided the JOLs. The children were reminded that they would be asked to take a test, on which they had to show that they had learned the meaning of the concepts. They were instructed to indicate with a mark those concepts they wished to restudy. However, participants were not given the opportunity to restudy the concepts they had selected because the aim of the study was to examine the effect of prior knowledge and JOLs on item selection.

After selecting items for restudy participants received the recall test (the fifth phase), they were asked to write down the correct meaning of the previously studied concepts.

The sixth phase followed this test, the participants were asked to review their test responses in order to provide SSJs about the correctness of their answer. Participants were not able to compare

their answers to a standard when self-scoring their test responses. SSJs were provided by marking the squares ranging from incorrect to fully correct which were depicted next to the tested concepts.

#### 2.1.4. Data analysis

Types of responses measured in the study include: (a) pretest responses, which were scored on an interval level ranging from 0 (incorrect) to 1 (fully correct), (b) JOLs, which were measured on an 11-point interval scale ranging from 0 to 10 (JOLs were then transformed to a 0–1 scale in order to make them comparable to the other outcome measures), (c) restudy selections, which were measured on a binary scale (0 = concept was not selected for restudy; 1 = concept was selected), and (d) SSJs, which were scored on an interval scale ranging from 0 (incorrect) to 1 (fully correct).



Each studied concept consisted of 2–4 idea units ( $M = 3.0$ ,  $SD = .64$ ), which was comparable to the concepts used by Dunlosky et al. (2005). Credit was given to the responses in line with the scoring procedure used by Dunlosky et al. (2005) and Rawson and Dunlosky (2007). Responses were scored using a gist criterion that is, the idea was scored if it was either provided verbatim or as a paraphrase of the idea unit. Pretest (prior knowledge) responses were rated as omissions (no response), inaccurate (a completely incorrect response, containing no idea units of the meaning of the concept), partially correct (a response contained at least one idea unit), or fully correct (a response containing all the main idea units). Two independent raters scored 43% of the pretest responses, and showed high inter-rater reliability (Cohen's Kappa = .90). Agreement was reached about the idea units where scoring diverged. As a result of the high inter-rater reliability, one rater scored the remaining pretest responses. Posttest (recall) responses were scored as omission errors (no response was given, no credit), commission errors (a completely incorrect response was given, no credit), partially correct responses (a response contained at least one idea unit, partial credit) and fully correct responses (containing all the idea units, full credit). For example, participants studied the concept *Cider* (concept) – *an alcoholic drink made of apples* (meaning). Then the participants were asked to provide the meaning of the concept *cider* at the posttest. When the response was: *A drink with alcohol*, this was scored as partially correct because it contains two of the three idea units. When participants responded with: *A drink made of apples which contains alcohol*, this response was scored as being fully correct. A response containing none of the idea units was scored as a commission error, such as answering for *cider*: *somebody who is working in prison*. Two raters scored 18% of the posttest responses, Kappa = .82. Agreement was reached about the idea units where scoring diverged and one rater subsequently scored the remaining pretest responses.

Consistent with previous studies (Nelson, 1984; Nelson & Dunlosky, 1991), JOL accuracy was measured by intra-individual gamma correlations between JOLs and test performance on the items (the closer to +1, the more accurate). The gamma correlation measures whether learners can judge their understanding of each concept relative to other concepts. Concordance occurs when the meaning of a concept with a JOL that is higher than that of another item is recalled, while the meaning of that second concept is not recalled. Discordance occurs when the meaning of a concept with a JOL that is higher of that of another concept is not recalled, while the meaning of that second concept is recalled. The gamma correlation is computed as follows:  $(\text{Concordances} - \text{Discordances}) / (\text{Concordances} + \text{Discordances})$ . Refer to Nelson (1984) for a discussion why a gamma correlation is the best measure of relative accuracy, and more preferable than a Spearman or Pearson correlation.

Restudy accuracy was measured by the intra-individual gamma correlations between JOLs and restudy choice (as a person should select items for restudy for which he provided low JOLs, a correlation closer to –1 indicates more accurate restudy selections, Thiede & Dunlosky, 1999). For seven participants, JOL accuracy could not be calculated because of invariance in either the JOLs or the test score. Restudy accuracy could not be calculated for 11 children due to invariance in selection of items for restudy or invariance in their test score.

Calibration accuracy of SSJs was measured per item as the difference between the SSJ and the objective score for the recall response. Calibration accuracy increases as the absolute difference between SSJs and performance decreases. With the calibration score, the directionality between judgments and actual performance is taken into account, so this measure indicates the extent to which one is overconfident or underconfident. As SSJs and test

**Table 1**

Pretest responses (percentage and total amount of omissions, inaccurate, partially correct, and fully correct) for the two grade cohorts.

Grade level	Prior knowledge: omission	Prior knowledge: inaccurate	Prior knowledge: partially correct	Prior knowledge: fully correct
Grade 3 and 4	74% (1452)	14.8% (290)	10.2% (201)	.9% (18)
Grade 5 and 6	72.2% (1382)	14.7% (279)	9.6% (183)	2.9% (56)

scores were both rated as 0, .5, or 1, calibration scores range from –1 to +1. A calibration score of 0 indicates perfect calibration, scores below 0 indicate underconfidence and scores above 0 indicate overconfidence.

Because of the hierarchical structure of the data (items were nested in participants, and participants were nested in classrooms and grade levels), the effect of inaccurate prior knowledge on commission errors, metacognitive judgments, and on calibration was investigated with multilevel regression analyses conducted in SPSS 19.

## 2.2. Results

### 2.2.1. Descriptive statistics

Participants activated prior knowledge on 27% of the items at the pretest, these responses were scored as inaccurate, partially correct, or fully correct. The remaining items (73%) were left blank by the participants and these were scored as omissions. Table 1 shows descriptive statistics regarding the percentage and the total amount prior knowledge activation (omissions, inaccurate, partially correct, or fully correct). At the recall test, participants provided responses on 44% of the items; the remaining 56% were scored as omission errors. Table 2 shows the percentage and total amount of omission errors, commission errors, partially correct, and fully correct responses.

Table 3 presents descriptive statistics about Mean JOLs, the percentage of concepts selected for restudy, credit for recall test responses, and Mean SSJs following the four categories of prior knowledge activation (omissions, inaccurate, partially correct, and fully correct). Table 4 presents descriptive statistics about credit for prior knowledge, Mean JOLs, the percentage of concepts selected for restudy, and Mean SSJs in relation to the four categories of recall test responses (omission errors, commission errors, partially correct responses, and fully correct responses).

As would be expected given the delayed JOL procedure, JOL accuracy was quite high,  $M$  gamma correlation = .78 ( $SD = .39$ ). The strength of the gamma correlation for restudy accuracy,  $M$  gamma correlation = –.80 ( $SD = .39$ ) shows that children selected concepts for restudy that they gave low JOLs and were less likely to select concepts for which they provided high JOLs.

Fig. 2 shows the calibration scores for the four different categories of the recall responses (omission errors, commission errors, partially correct, fully correct) for both age cohorts. For the third- and fourth-grade students, mean posttest self-score judgments for commission errors were .72 ( $SD = .24$ ), indicating high

**Table 2**

Recall test responses (percentage and total amount of omission errors, commission errors, partially correct, and fully correct) for the two grade cohorts.

Grade level	Recall test: omission error	Recall test: commission error	Recall test: partially correct	Recall test: fully correct
Grade 3 and 4	56.3% (1108)	12.6% (248)	19.6% (386)	11.5% (226)
Grade 5 and 6	39.9% (739)	9.1% (168)	21.4% (396)	29.6% (549)

**Table 3**

Mean JOLs, percentage of concepts selected for restudy, mean recall response credit, and mean SSJs of 3rd/4th and 5th/6th grade cohorts following omitted, inaccurate, partially correct and fully correct prior knowledge activation (SD of the mean in parentheses).

Grade level	Prior knowledge: omission	Prior knowledge: inaccurate	Prior knowledge: partially correct	Prior knowledge: fully correct
<i>JOLs (range 0–1)</i>				
Grade 3 and 4	.42 (.41)	.78 (.32)	.94 (.18)	.99 (.03)
Grade 5 and 6	.47 (.41)	.8 (.31)	.91 (.18)	.96 (.1)
<i>Restudy selections (%)</i>				
Grade 3 and 4	60% (.49)	28% (.45)	12% (.33)	6% (.23)
Grade 5 and 6	51% (.41)	24% (.43)	10% (.29)	11% (.31)
<i>Recall response credit (range 0–1)</i>				
Grade 3 and 4	.15 (.31)	.26 (.36)	.56 (.31)	.72 (.35)
Grade 5 and 6	.33 (.43)	.50 (.41)	.68 (.27)	.82 (.32)
<i>SSJs (range 0–1)</i>				
Grade 3 and 4	.24 (.40)	.59 (.41)	.81 (.31)	.92 (.19)
Grade 5 and 6	.39 (.45)	.72 (.37)	.88 (.22)	.50 (.46)

overconfidence in these errors. Children often expected to receive credit for responses that were entirely incorrect. The children expected to receive no credit for only 5.3% of their commission errors, whereas they expected to receive partial credit for 42.3% and full credit for 52.4% of the commission errors.

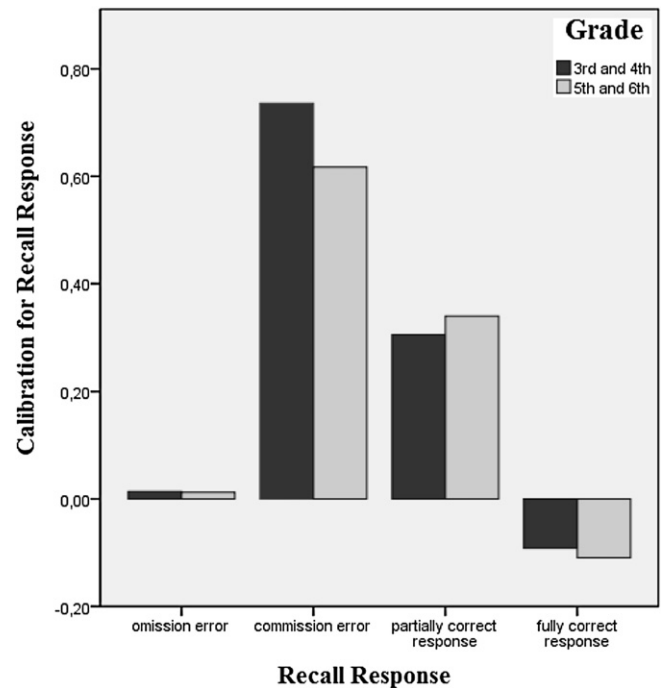
### 2.2.2. Relation between inaccurate prior knowledge activation and recall responses

Table 5 shows the effect of the four categories of pretest prior knowledge activation (omission, inaccurate, partially correct, fully correct) on the percentage of recall test responses (omission error, commission error, partially correct, fully correct). These results show that when participants activated inaccurate prior knowledge on the pretest, this resulted in an omission on the recall test for 25.1% of the responses, in a commission error for 36.6% of the responses, in partially correct for 25.4%, and in a fully correct for 12.9% of the responses. A binary logistic multilevel regression analysis, (0 = no ability to activate any prior knowledge, 1 = inaccurate prior knowledge) was used to investigate the effect of prior knowledge activation at the pretest

**Table 4**

Mean pretest response credit, mean JOLs, percentage of concepts selected for restudy, and mean SSJs of 3rd/4th and 5th/6th grade cohorts for omission errors, commission errors, partially correct responses and fully correct responses at the recall test (SD of the mean in parentheses).

Grade level	Recall test: omission error	Recall test: commission error	Recall test: partially correct response	Recall test: fully correct response
<i>Pretest response credit (range 0–1)</i>				
Grade 3 and 4	.01 (.05)	.03 (.14)	.17 (.25)	.16 (.28)
Grade 5 and 6	0 (.04)	.03 (.17)	.16 (.26)	.14 (.3)
<i>JOLs (range 0–1)</i>				
Grade 3 and 4	.26 (.33)	.84 (.28)	.89 (.23)	.89 (.23)
Grade 5 and 6	.19 (.27)	.69 (.35)	.85 (.26)	.85 (.26)
<i>Restudy selections (%)</i>				
Grade 3 and 4	76% (.43)	19% (.39)	15% (.36)	12% (.33)
Grade 5 and 6	71% (.46)	35% (.48)	19% (.40)	22% (.41)
<i>SSJs (range 0–1)</i>				
Grade 3 and 4	.01 (.1)	.74 (.3)	.81 (.29)	.91 (.23)
Grade 5 and 6	.01 (.11)	.62 (.36)	.84 (.25)	.89 (.22)



**Fig. 2.** Calibration of self-score judgments for recall responses (omission errors, commission errors, partially correct, and fully correct) for the two grade cohorts. Calibration is measured as the difference between self-score judgments and objective credit for recall responses. Positive values indicate overconfidence, negative values underconfidence.

on commission errors at the recall test. When inaccurate prior knowledge was activated, this led to more commission errors at the recall test than when learners were not able to activate any prior knowledge (Hypothesis 1), Wald Chi-Square = 1, 109.47,  $p = .0001$  ( $b = 1.83$ ,  $SE = .17$ , Wald CI = 1.49–2.17). When inaccurate prior knowledge was activated at the pretest, 36.6% of this inaccurate prior knowledge became a commission error at the recall test. When learners were not able to activate any prior knowledge, only 8.5% of these omissions became a commission error at the recall test.

Table 5 shows that inaccurate prior knowledge not only led to more commission errors, but also to more partially correct responses than when learners were not able to activate any prior knowledge, Wald Chi-Square = 1, 30.62,  $p < .0001$  ( $b = .86$ ,  $SE = .16$ , Wald CI = .56–1.17). In addition, Table 5 shows that when participants were not able to activate any prior knowledge at the pretest, these concepts remained omission errors at the recall test significantly more often than when inaccurate prior knowledge was activated, Wald Chi-Square = 1, 39.38,  $p < .0001$  ( $b = 1.09$ ,  $SE = .17$ , Wald CI = .75–1.44).

### 2.2.3. Effect of inaccurate prior knowledge activation on monitoring and regulation

A mixed model regression analysis was conducted to investigate whether JOLs were inappropriately high after activation of inaccurate prior knowledge, when compared to no activation of prior knowledge at all (Hypothesis 2a). Items were included as repeated measures, and a random intercept was included for every participant. Only the JOLs provided for concepts for which the recall responses would not receive credit (omissions and commission errors) were used in this analysis. For these incorrect responses, accurate JOLs should be low. Regression coefficients, standard errors, and 95% confidence intervals are reported for inaccurate prior knowledge activation.

**Table 5**

Percentage of students' recall test omission and commission errors, and partially and fully correct responses following omitted, inaccurate, partially correct and fully correct prior knowledge activation. Findings are tabled for both grade cohorts.

	Recall test responses			
	Test: omission error	Test: commission error	Test: partially correct response	Test: fully correct response
<i>Grade 3 and 4</i>				
Prior knowledge: omission	70.1%	8.5%	12.6%	8.8%
Prior knowledge: inaccurate	25.1%	36.6%	25.4%	12.9%
Prior knowledge: partially correct	7.5%	6.0%	60.8%	25.6%
Prior knowledge: fully correct	0%	11.1%	33.3%	55.6%
<i>Grade 5 and 6</i>				
Prior knowledge: omission	52.1%	7.7%	14.5%	25.8%
Prior knowledge: inaccurate	13.2%	20.9%	31.9%	34.1%
Prior knowledge: partially correct	.6%	2.3%	58.8%	38.4%
Prior knowledge: fully correct	1.8%	7.1%	17.9%	73.2%

Results show that JOLs for concepts for which participants did not receive credit at the test were significantly higher after activation of inaccurate prior knowledge than after no activation of prior knowledge,  $t(1274.5) = 15.91$ ,  $p < .0001$  ( $b = 4.47$ ,  $SE = .28$ ,  $CI = 3.92\text{--}5.02$ ). JOLs (measured on a scale ranging from 0 to 10) for incorrect recall test responses were on average 7.3 ( $SD = 3.6$ ) after activation of inaccurate knowledge, and 2.9 ( $SD = 3.6$ ) after no activation of prior knowledge.

Fig. 3 depicts calibration accuracy of SSJs for recall test responses following omissions, inaccurate, partially, and fully correct prior knowledge activation at the pretest. A mixed model regression analysis was used to investigate whether inaccurate prior knowledge activation led to more overconfidence compared to not being able to activate any prior knowledge (Hypothesis 2b). Items were included as repeated measurement and a random intercept was included for the participants. Calibration accuracy for the SSJs showed that children were more overconfident when judging the

quality of their recall after inaccurate prior knowledge activation at the pretest ( $M = .32$ ,  $SD = .42$ ) than when they were unable to activate prior knowledge ( $M = .09$ ,  $SD = .28$ ),  $t(1657) = 11.41$ ,  $p < .0001$  ( $b = .23$ ,  $SE = .02$ ,  $CI = .19\text{--}.27$ ).

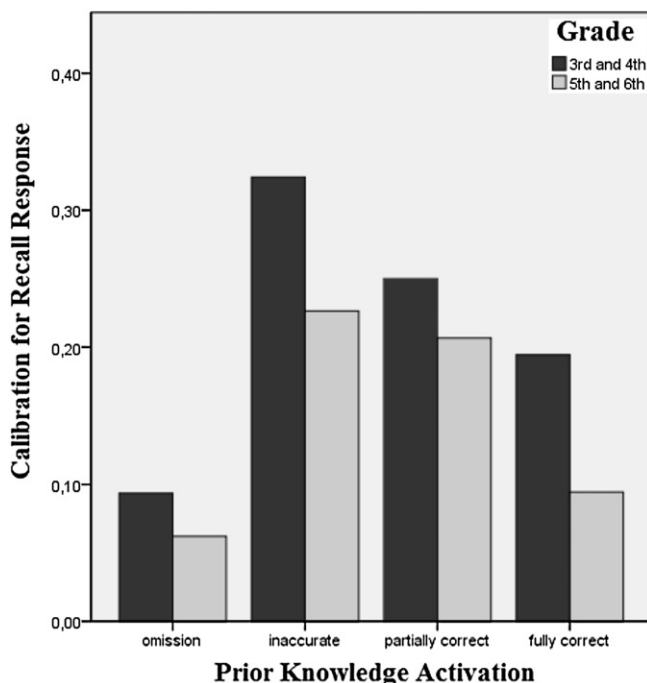
Because the selection of items for restudy was measured on a binary scale (0 = item not selected, 1 = item selected), a generalized linear model for binary data was used to investigate the effect of prior knowledge activation on item selection. When accurately selecting items for restudy, participants would choose concepts of which they do not recall the meaning. Therefore, this analysis was again conducted on the concepts for which the participants would receive no credit at the recall test. We investigated whether activation of inaccurate prior knowledge at the pretest would lead children to more often discard these concepts from further study when compared to no activation of prior knowledge (Hypothesis 3). After activation of inaccurate prior knowledge at the pretest, concepts for which they would not receive credit at the recall test were less often selected for restudy, than after no activation of prior knowledge. After activation of inaccurate prior knowledge, 34% of the concepts was selected ( $SD = .48$ ) versus 75% of the concepts ( $SD = .45$ ) when no prior knowledge could be activated. This difference was significant, Wald Chi-Square = 1, 55.23,  $p < .0001$  ( $b = -1.53$ ,  $SE = .20$ , Wald  $CI = -1.93$  to  $-1.12$ ).

### 2.3. Discussion

The results of Study 1 showed that third- and fourth-grade children were highly overconfident when they evaluated the quality of their commission errors. This finding is consistent with the results of Lipko, Dunlosky, Hartwig, et al. (2009). The results showed that commission errors on the recall test occurred more often after activation of inaccurate prior knowledge at the pretest, than after no activation of prior knowledge, or after (partially) correct activation (Hypothesis 1). A substantial part of the commission errors emanated from inaccurate prior knowledge at the pretest.

When providing JOLs for concepts that were not yet understood, children's JOLs were higher after activation of inaccurate prior knowledge than after not being able to activate any prior knowledge (Hypothesis 2a). Moreover, the results showed that activation of inaccurate prior knowledge was detrimental to calibration accuracy of the SSJs (Hypothesis 2b). That is, children were more overconfident after commission errors than after omissions or (partially) correct prior knowledge activation.

With regard to regulation, children more often prematurely discarded concepts for which they activated inaccurate prior knowledge, whereas they more often decided to restudy the



**Fig. 3.** Calibration of self-score judgments for recall responses following prior knowledge activation (omissions, inaccurate, partially correct, and fully correct) for the two grade cohorts. Calibration is measured as the difference between self-score judgments and objective credit for recall responses.



concepts for which they were not able to activate any prior knowledge (Hypothesis 3).

In sum, our findings indicate that commission errors, and overconfidence in these errors, in part results from the activation of inaccurate prior knowledge. Possibly, inaccurate prior knowledge activation leads to inaccurate JOLs and SSJs because learners base these judgments on accessibility of information, instead of the quality of the accessible information (Koriat, 1993). However, the present findings were obtained for third- and fourth-grade students. At this young age, metacognitive skills are still developing. There is evidence that between the ages of 9 and 12, metacognitive skills become more accurate and learners become less overconfident (e.g. Flavell, 1970; Krebs & Roebers, 2010, 2011; Schneider & Pressley, 1997). Therefore, it might be that inaccurate activation of prior knowledge has smaller effects on performance, metacognitive judgments, and calibration for older primary-school children. To test this possibility, we conducted a second study with fifth- and sixth-grade students. The same hypotheses were tested as in the first study, and the same procedure was used; only the concepts in the study task differed to make them age appropriate.

### 3. Study 2

#### 3.1. Method

##### 3.1.1. Participants

Participants were 95 primary-school children (48.4% male, 51.6% female,  $M = 10.6$  years,  $SD = .8$ ) from two fifth- ( $n = 48$ ) and two sixth-grade ( $n = 47$ ) classrooms from the same Dutch primary-school as participants in Study 1. Study 2 did not include any of the Study 1 participants because both studies occurred in the same school year. All participants were fluent enough in Dutch to follow school lessons and to understand our instructions.

##### 3.1.2. Materials and procedure

The procedure of Study 2 was exactly the same as that of the first study, except for the concepts. Based on the measure of lexical richness (Vermeer, 2000) and a pilot study with four participants, Dutch concepts were selected in the same manner as in the first study.

##### 3.1.3. Data analysis

Studied concepts consisted of two to four idea units ( $M = 3.0$ ,  $SD = .71$ ). The prior knowledge responses and recall responses were scored according to the same criteria as used in Study 1. Two independent raters scored 42% of the pretest and 13% of the recall test responses and showed a high level of agreement ( $Kappa = .90$  and  $.79$  respectively). Scoring of the first rater was used and this rater scored the remaining pretest and recall test responses.

Relative accuracy of JOLs could not be calculated for two participants due to invariance in JOL or test score. Gamma correlations indicating the efficiency of regulation of study could not be calculated for three participants due to invariance in selection or test score.

#### 3.2. Results

##### 3.2.1. Descriptive statistics

Participants provided a prior knowledge response at the pretest for 27% of the items. Table 1 shows the percentage and total amount of omissions, inaccurate, partially, and fully correct prior knowledge for the provided pretest responses. At the recall test, participants provided responses for 60% of the items. Table 2 shows the percentage and total amount of omission errors, commission errors, partially correct, and fully correct responses at the recall test.

Descriptive statistics in relation to the four categories of prior knowledge activation (omissions, inaccurate, partially correct, fully correct) and the four categories of recall responses (omission errors, commission errors, partially correct responses, fully correct responses) are presented in Tables 3 and 4, respectively. Because different materials were used for the two age groups we could not test whether there were significant differences between the younger and the older cohort.

As in Study 1, relative JOL accuracy was high,  $M$  gamma correlation =  $.76$  ( $SD = .27$ ), and the gamma correlation for regulation of study,  $\gamma = -.81$  ( $SD = .40$ ), indicated that participants used their JOLs to select concepts for restudy.

Calibration accuracy for the four categories of recall responses (omission errors, commission errors, partially correct responses, and fully correct responses) is depicted in Fig. 2. The mean of the SSJs for commission errors at the test was  $.62$  ( $SD = .26$ ), indicating that participants were overconfident when judging the quality of these recall responses. For 16% of these commission errors they expected to receive no credit, for 44.4% they expected to receive partial credit, and for 39.5% they expected to receive full credit.

##### 3.2.2. Relation between inaccurate prior knowledge activation and recall responses

Table 5 displays the relation between pretest prior knowledge activation and responses at the recall test. As the table shows, when participants activated inaccurate prior knowledge on the pretest, there was an omission on the recall test for 13.2% of the responses, a commission error for 20.9%, a partially correct response on the recall test for 31.9% of the responses, and a fully correct response on the recall test for 34.1% of the responses. When participants activated inaccurate prior knowledge, they were more likely to make commission errors at the test than when they were unable to activate prior knowledge, 20.9% compared with 7.7%, (Hypothesis 1), Wald Chi-Square = 1, 37.87,  $p < .0001$  ( $b = 1.16$ ,  $SE = .19$ , Wald CI =  $.79$ – $1.53$ ). In addition, activation of inaccurate prior knowledge led more often to partially correct responses at the recall test than when participants were unable to activate prior knowledge, Wald Chi-Square = 1, 45.63,  $p < .0001$  ( $b = 1.02$ ,  $SE = .15$ , Wald CI =  $.72$ – $1.31$ ). When participants were unable to activate prior knowledge, this significantly more often remained an omission error at the recall test than when inaccurate prior knowledge was activated, Wald Chi-Square = 1, 151.34,  $p < .0001$  ( $b = 1.86$ ,  $SE = .15$ , Wald CI =  $1.58$ – $2.19$ ).

##### 3.2.3. Effect of inaccurate prior knowledge activation on monitoring and regulation

A mixed model regression analysis showed that JOLs for the incorrect recall responses were significantly higher after pretest activation of inaccurate prior knowledge than after no activation of prior knowledge,  $t(818,79) = 10.38$ ,  $p < .0001$  ( $b = 3.37$ ,  $SE = .32$ , CI =  $2.73$ – $4.00$ ). JOLs for incorrect recall responses were on average 6.1 ( $SD = 2.4$ ) after activation of inaccurate prior knowledge, and 2.4 ( $SD = 3.1$ ) when learners were unable to activate prior knowledge (Hypothesis 2a).

Fig. 3 depicts calibration accuracy following prior knowledge activation. The influence of inaccurate prior knowledge activation on posttest calibration accuracy of SSJs was investigated with a mixed model regression analysis (Hypothesis 2b). Calibration scores indicated more overconfidence after inaccurate prior knowledge activation ( $M = .23$ ,  $SD = .39$ ) than after no activation ( $M = .06$ ,  $SD = .28$ ),  $t(1578) = 8.14$ ,  $p < .0001$  ( $b = .16$ ,  $SE = .02$ , CI =  $.12$ – $.20$ ).

To investigate the effect of inaccurate prior knowledge on selection of items for restudy for which incorrect recall would follow, a generalized linear model for binary data was used

(Hypothesis 3). Findings showed that after activation of inaccurate prior knowledge items were more often prematurely discarded from study (only 38% of the concepts were selected for restudy,  $SD = 4.8$ ) than when participants were unable to activate prior knowledge (67% of the concepts were selected,  $SD = 4.9$ ), Wald Chi-Square = 1, 28.62,  $p < .0001$  ( $b = -1.23$ ,  $SE = .23$ , Wald CI =  $-1.68$  to  $-.78$ ).

### 3.3. Discussion

The results of Study 2 show the same pattern as those of Study 1. Even though the older children had a lower frequency of commission errors, and their absolute level of overconfidence was lower than those of the younger children, overconfidence in commission errors was high. Again, activation of inaccurate prior knowledge before study led to inappropriately high JOLs, less accurate calibration, and less efficient selection for restudy than when children were not able to activate any prior knowledge.

## 4. General discussion

The aim of the present study was to investigate the effect of inaccurate prior knowledge activation on young learners' commission errors and their overconfidence in these errors when learning concepts. To the best of our knowledge, this study is the first to investigate the effect of inaccurate prior knowledge activation on children's calibration accuracy.

The study shows that inaccurate prior knowledge negatively affects children's metacognitive judgments in several ways when learning novel concepts. First of all, children did not accurately monitor the quality of information that was accessible in their memory. The findings showed that JOLs prior to recall responses for which they would not receive any credit were inappropriately high after activation of inaccurate prior knowledge (Hypothesis 2a). Moreover, results showed that activation of inaccurate prior knowledge was not only detrimental for monitoring judgments during learning, but also for calibration after test taking. That is, children were overconfident when self-scoring the quality of their recalled ideas after activation of inaccurate prior knowledge (Hypothesis 2b). Finally, this also applied when selecting concepts for further study. Children often prematurely discarded concepts for which they did not accurately recall the meaning after activation of inaccurate prior knowledge (Hypothesis 3).

Our findings show that when inaccurate prior knowledge was activated at the pretest, this more often resulted in a commission error at the recall test when compared with concepts for which no prior knowledge could be activated (Hypothesis 1). This implies that when activated prior knowledge is inaccurate, children are often not aware of this. Presumably, learners focus on the mere accessibility of information in memory instead of on the quality (Koriat, 1993), and therefore their inaccurate prior knowledge is often not corrected through further learning.

Note that when children *did* activate inaccurate prior knowledge at the pretest, recall test responses were more often partially correct than when children were not able to activate any prior knowledge (in that case pretest responses mainly remained omissions). At first sight, this seems to make the interpretation that inaccurate prior knowledge deteriorates self-regulated learning somewhat paradoxical. However, the problematic finding is that when learners activated inaccurate prior knowledge for concepts, later test responses not only became more often partially correct, but also more often turned into commission errors (on average 29% versus 8% when learners could not activate

any prior knowledge). Calibration was most inaccurate for these commission errors: When children judged the quality of these errors they expected to receive credit for these completely incorrect responses. Moreover, findings show that it is unlikely that children will decide to allocate their study time to their commission errors: On average participants selected only 37% of the commission errors for restudy (whereas they decided to restudy 74% of the concepts for which they made omission errors). Our findings imply that when children engage in self-regulated learning in a naturalistic setting, they mainly allocate their study time to omissions, and often decide not to further study commission errors. It thus seems likely that recall performance will improve for omissions, but it is less likely that recall performance will improve for commission errors. In sum, our findings show that inaccurate prior knowledge activation can lead to commission errors, and inaccurate prior knowledge seems therefore more problematic for calibration and self-regulated learning than when learners do not have any prior knowledge accessible prior to studying.

It should be noted that the percentage of items that were responded to on the pretest, and as a consequence, the percentage of inaccurately activated prior knowledge, was relatively low in both studies. At the pretest as well as the recall test participants were allowed omissions (i.e., not providing any response), because we would have artificially inflated the frequency of inaccurate responses on the pretest if we would have forced the children to provide a response. The low percentage of responses in general confirms that the concepts we selected were largely unknown to the participants (a necessity because otherwise there would not be much left to learn).

In the present study, learners were asked to attempt to activate prior knowledge at a pretest prior to study. This may have been a limitation of the study because in real-life learners are usually not confronted with a pretest. However, it is likely that inaccurate prior knowledge also hinders self-regulated learning when learners are not explicitly asked to activate prior knowledge at a pretest, as other research has shown that learners base their monitoring judgments on their assessment of prior knowledge (Dinsmore & Parkinson, 2013).

Relative accuracy of the JOLs was high, in line with prior findings on delayed JOLs (Nelson & Dunlosky, 1991; Rhodes & Tauber, 2011). In addition, consistent with previous research (De Bruin et al., 2011; Metcalfe & Finn, 2008), the children decided to restudy the concepts for which they provided low JOLs. This indicates that children can accurately discriminate between concepts on the basis of accessibility of these concepts in memory as a cue, such as is measured by the correlations between JOLs and performance. However, our findings reveal that JOLs and restudy selections were not always accurate. When children are asked to judge the *quality* of these ideas, which is required when making absolute judgments, their monitoring was very poor. The monitoring judgments were inappropriately high when children had inaccurate information accessible, and children often prematurely discarded these items from restudy.

Even though we were unable to statistically compare the age cohorts, the equal pretest scores indicate that these materials were of similar difficulty. The results on activation of inaccurate prior knowledge between cohorts also largely parallel each other. It seems that learners in the older age cohort were able to overwrite their inaccurate prior knowledge more often. They made less commission errors at the recall test, and showed less overconfidence in these errors than the younger age cohort. Consistent with Krebs and Roebbers (2010, 2011), the older age group was better calibrated than the younger one. However, overconfidence occurred for commission errors for the older age group too, and

inaccurate prior knowledge led to inflated JOLs, less efficient regulation of study, and less accurate calibration.

The focus of the present study was on calibration for omissions and commission errors, because to improve learning it is most important to investigate monitoring and control for incorrect responses (Krebs & Roebers, 2010). Therefore, we did not explicitly focus on evaluations of the partially correct responses that contained some correct ideas but were incomplete. However, even though we did not specifically focus on partially correct responses, it has to be noted that the children also overestimated the quality of their recall when evaluating these responses. Previous studies have also shown that overconfidence is not only a problem with commission errors. When judging the quality of recall of partially correct responses, learners often expect to receive full credit (Dunlosky & Lipko, 2007; Dunlosky et al., 2005). The present results indicate that the absolute level of calibration accuracy was not only poor after inaccurate prior knowledge activation, but also after activation of partially correct prior knowledge. Therefore, it appears that partially correct prior knowledge can also be hard to correct through instruction. Future research should further investigate the effect of activation of partially correct prior knowledge on calibration and self-regulated learning.

The present study shows that inaccurate prior knowledge can contribute to commission errors at the recall test, and to poor calibration for these errors. These findings reinforce the need for future research to investigate interventions which improve children's self-regulated learning, by focusing not only on accessibility but also on the quality of information in memory. Even though learners did overwrite some of their inaccurate prior knowledge in our studies, the findings demonstrate that this does not necessarily happen, especially not for the 8/9-year-old children.

Unfortunately, learners often have problems overwriting their inaccurate knowledge (e.g., Chi, Slotta, & De Leeuw, 1994). Research has shown that to overwrite accessible inaccurate prior knowledge, it is necessary that learners co-activate this together with the correct information, and that they become aware that the two conflict with each other (Van den Broek & Kendeou, 2008). It is possible that an explicit feedback intervention after prior knowledge activation might be promising to improve calibration accuracy. The definitions of the concepts were provided in the learning phase, and could have implicitly served as feedback, but apparently this was not sufficient. Explicit feedback could make learners aware of the fact that their prior knowledge conflicts with the information presented in the learning phase. Metcalfe and Finn (2012) found that although children's errors were not self-corrected without feedback, when feedback was given about errors performance improved drastically.

To conclude, as Ausubel, Novak, and Hanesian (1978) stated: "The most important factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly" (p. 163). Our findings add to this statement that what the learner *believes* to know also influences his learning, not only directly, but also indirectly by affecting monitoring and regulation of learning. Therefore, the identification of what children already know, and the detection of inaccuracies in their prior knowledge, might be important to improve their self-regulated learning.

## Acknowledgment

This research was supported by the Netherlands Institute for Scientific Research (NWO), grant 411.07.151. We would like to thank Eddy Bakker, Heleen van der Ham, Renske van der Hoeven, Jimmie Leppink, and Andrea Oudkerk Pool for help with data analysis, and Bill Wrigley for comments on a previous draft of this manuscript.

## References

- Alexander, P. A., & Jetton, T. L. (2000). Learning from text: a multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research, Vol. III* (pp. 285–310). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ausubel, D., Novak, J., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2nd ed.). New York: Holt, Rinehart & Winston.
- Barnett, J., & Hixon, J. (1997). A developmental study of self-assessment in elementary school children. *Journal of Educational Research*, 90, 170–174.
- Chi, M. T. H., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: a theory of conceptual change for learning science concepts. *Learning and Instruction*, 4, 27–43. [http://dx.doi.org/10.1016/0959-4752\(94\)90017-5](http://dx.doi.org/10.1016/0959-4752(94)90017-5).
- De Bruin, A. B. H., Thiede, K. W., Camp, G., & Redford, J. (2011). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology*, 109, 294–310. <http://dx.doi.org/10.1016/j.jecp.2011.02.005>.
- Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction*, 24, 4–14.
- Dochy, F., Segers, M., & Buehl, M. (1999). The relation between assessment practices and outcomes of studies: the case of research on prior knowledge. *Review of Educational Research*, 69, 147–188. <http://dx.doi.org/10.2307/1170673>.
- Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: differential and sufficient aspects. *Journal of Experimental Child Psychology*, 47, 274–296. [http://dx.doi.org/10.1016/0022-0965\(89\)90033-7](http://dx.doi.org/10.1016/0022-0965(89)90033-7).
- Dunlosky, J., Hartwig, M. K., Rawson, K. A., & Lipko, A. R. (2011). Improving college students' evaluation of text learning using idea-unit standards. *Quarterly Journal of Experimental Psychology*, 64, 467–484. <http://dx.doi.org/10.1080/17470218.2010.502239>.
- Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: a brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16, 228–232. <http://dx.doi.org/10.1111/j.1467-8721.2007.00509.x>.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, 22(4), 271–280. <http://dx.doi.org/10.1016/j.learninstruc.2011.08.003>.
- Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, 52, 551–565. <http://dx.doi.org/10.1016/j.jml.2005.01.011>.
- Flavell, J. H. (1970). The development of mediated memory. In H. W. Reese, & L. P. Lipsitt (Eds.), *Advances in child development and behavior*, Vol. 5. New York: Academic Press.
- Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: a focus on calibration. In J. Dunlosky, & R. Bjork (Eds.), *Handbook of memory and metacognition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: examining the nature of judgments of confidence. *Learning and Instruction*, 24, 37–47.
- Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition & Learning*, 4, 161–176. <http://dx.doi.org/10.1007/s11409-009-9042-8>.
- Koriat, A. (1993). How do we know what we know: the accessibility model of the feeling of knowing. *Psychological Review*, 100, 609–639. <http://dx.doi.org/10.1037/0033-295X.100.4.609>.
- Koriat, A. (1995). Dissociating knowing and the feeling of knowing: further evidence for the accessibility model. *Journal of Experimental Psychology: General*, 124, 311–333. <http://dx.doi.org/10.1037/0096-3445.124.3.311>.
- 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. <http://dx.doi.org/10.1037/0096-3445.126.4.349>.
- Koriat, A. (2011). Subjective confidence in perceptual processes: a test of the self-consistency model. *Journal of Experimental Psychology: General*, 140, 117–139. <http://dx.doi.org/10.1037/a0022171>.
- Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and the accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 34–53. <http://dx.doi.org/10.1037/0278-7393.27.1.34>.
- Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy. In P. Chambres, M. Izaute, & P. J. Maraescaud (Eds.), *Metacognition: Process, function and use* (pp. 1–17). Boston: Kluwer Academic Publishers.
- Krebs, S. S., & Roebers, C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology*, 80, 325–340. <http://dx.doi.org/10.1348/000709910X485719>.
- Krebs, S. S., & Roebers, C. M. (2011). The impact of retrieval processes, age, general achievement level, and test scoring scheme for children's metacognitive monitoring and controlling. *Metacognition and Learning*, . <http://dx.doi.org/10.1007/s11409-011-9079-3>.
- Lipko, A. R., Dunlosky, J., Hartwig, M. K., Rawson, K. A., Swan, D., & Cook, D. (2009). Using standards to improve middle school students' accuracy at evaluating the quality of their recall. *Journal of Experimental Psychology: Applied*, 15, 307–318. <http://dx.doi.org/10.1037/a0017599>.
- 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, 152–166. <http://dx.doi.org/10.1016/j.jecp.2008.10.002>.
- Lipson, M. Y. (1982). Learning new information from text: the role of prior knowledge and reading ability. *Journal of Reading Behavior*, 14, 243–261.
- Maki, R. H., Shields, M., Wheeler, A., & Zaccilli, T. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, 97, 723–731. <http://dx.doi.org/10.1037/0022-0663.97.4.723>.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15, 174–179. <http://dx.doi.org/10.3758/pbr.15.1.174>.
- Metcalfe, J., & Finn, B. (2012). Hypercorrection of high confidence errors in children. *Learning & Instruction*, 22, 253–261. <http://dx.doi.org/10.1016/j.learninstruc.2011.10.004>.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109–133. <http://dx.doi.org/10.1037/0033-2909.95.1.109>.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: the "delayed-JOL effect". *Psychological Science*, 2, 267–270. <http://dx.doi.org/10.1111/j.1467-9280.1991.tb00147.x>.
- 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. <http://dx.doi.org/10.1111/j.1467-9280.1994.tb00502.x>.
- Nelson, T. O., & Narens, L. (1990). Metamemory: a theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 26 (pp. 125–171). New York: Academic Press.
- Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, 19, 228–242. <http://dx.doi.org/10.1016/j.learninstruc.2008.04.003>.
- Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, 19, 559–579. <http://dx.doi.org/10.1080/09541440701326022>.
- Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: a meta-analytic review. *Psychological Bulletin*, 137, 131–148. <http://dx.doi.org/10.1037/a0021705>.
- Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty*. Mahwah, NJ: Erlbaum.
- Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: evidence from a judgment-of-learning task. *Cognitive Development*, 15, 115–134. [http://dx.doi.org/10.1016/S0885-2014\(00\)00024-1](http://dx.doi.org/10.1016/S0885-2014(00)00024-1).
- Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415–429). New York: Routledge.
- Schraw, G., Kuch, F., & Gutierrez, A. P. (2013). Assessing the dimensionality of calibration measures used in monitoring research. *Learning and Instruction*, 24, 48–57.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of text. *Journal of Educational Psychology*, 95, 66–73. <http://dx.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, 1024–1037. <http://dx.doi.org/10.1037/0278-7393.25.4.1024>.
- Van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science text: the role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, 22, 335–351. <http://dx.doi.org/10.1002/acp.1418>.
- Verburg, M. (2008). *Van Dale woordenboek Nederlands voor VMBO & MBO: Het woordenboek in begrijpelijke taal*. Utrecht: Van Dale Lexicografie.
- Vermeer, A. (2000). Lexicale rijkdom, tekstmoeilijkheid en woordenschatgrootte. Beschrijving van de MLR, een woordenschat-analyse programma. *Toegepaste Taalwetenschap in Artikelen*, 64, 95–105.
- Wetzels, S. A. J., Kester, L., & Van Merriënboer, J. (2011). The influence of prior knowledge on the retrieval-directed function of note taking in prior knowledge activation. *British Journal of Educational Psychology*, 81, 274–291. <http://dx.doi.org/10.1348/000709910X517425>.