Studies in Educational Evaluation 58 (2018) 51–59

Contents lists available at ScienceDirect

Studies in Educational Evaluation

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

Accuracy of primary school children's immediate and delayed judgments of learning about problem-solving tasks

Martine Baars[a,*], Tamara van Gog[b], Anique de Bruin[c], Fred Paas[a,d]

[a] Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, The Netherlands
[b] Department of Education, Utrecht University, The Netherlands
[c] Department of Educational Development and Research, Maastricht University, The Netherlands
[d] Early Start Research Institute, University of Wollongong, Australia

ARTICLE INFO

Keywords:
Judgments of learning
Monitoring accuracy
Problem solving
Mental effort
Primary education

ABSTRACT

Accuracy of students' judgments of learning (JOLs) plays an important role in self-regulated learning. Most studies on JOL accuracy have focused on learning word pairs and text but problems-solving tasks are also very important in education. This study investigated whether children in grade 3 could differentiate in their JOLs between problem-solving tasks that varied in complexity. Participants ($N = 76$, 8–10 years old) engaged in solving four arithmetic problems, rated mental effort invested in each problem, gave either immediate or delayed JOLs, and completed a test containing isomorphic problems. The negative correlation that was found between invested mental effort and JOLs suggested that children's JOLs are sensitive to differences in complexity of the problem-solving tasks. Results on the relative and absolute accuracy of JOLs showed that immediate JOLs were numerically higher than delayed JOLs, and relative accuracy of immediate JOLs was moderately accurate, whereas delayed JOLs were not.

1. Introduction

Research has shown that monitoring accuracy, that is, accuracy of students' judgments of what information they have or have not yet learned, plays an important role in self-regulated learning. When these monitoring judgments are not accurate, students will not be able to make optimal study choices, for example about how they should allocate their study time and what information they need to restudy (Dunlosky & Lipko, 2007; Metcalfe, 2009). Research on ways of enhancing the accuracy of students' monitoring judgments has mainly focused on study materials consisting of paired associates, quizzes, or short expository texts (Bol, Hacker, O'Shea, & Allen, 2005; Hacker, Bol, & Bahbahani, 2008; Rhodes & Tauber, 2011; Thiede, Griffin, Wiley, & Redford, 2009). Much less is known about monitoring judgments regarding the kind of procedural problem-solving tasks typically seen in important school subject domains such as math or science (see Efklides, 2002).

Ackerman and Thompson (2014) described meta-reasoning as the process by which learners monitor and control reasoning, problem solving and decision-making processes. There are many different kinds of problem-solving tasks; they vary from insight problems to well-structured transformation problems that have a clearly defined goal and solution procedure, to ill-structured problems that do not have a well-defined goal or solution procedure. Well-structured problems, such as math and biology problems encountered in primary and secondary education, consist of a well-defined initial state, a known goal state, and can be solved using a constrained set of logical operators (Jonassen, 2011). Even though monitoring one's own performance and being able to regulate further learning is important when learning to solve problems, few studies have investigated these processes in problem-solving tasks. Here, we take a first step towards investigating whether primary school children differentiate in their monitoring judgments between math problem-solving tasks that differ in complexity, and by exploring the accuracy of immediate and delayed monitoring judgments.

1.1. Metacognition and self-regulated learning

Metacognition involves knowledge, monitoring, and control of a cognitive process, such as learning (Flavell, 1979; Serra & Metcalfe, 2009). Metacognition is held to play an important role in learning and especially self-regulated learning, because successful self-regulated learning depends on accurate monitoring and control processes (e.g., Winne & Hadwin, 1998). Research has shown that when metacognitive knowledge, monitoring and control are adequate, learning is enhanced

* Corresponding author at: Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands.
E-mail address: baars@essb.eur.nl (M. Baars).

https://doi.org/10.1016/j.stueduc.2018.05.010
Received 21 February 2018; Received in revised form 9 May 2018; Accepted 30 May 2018

(Azevedo & Cromley, 2004; Kornell & Metcalfe, 2006; Metcalfe, 2009; Thiede, Anderson, & Therriault, 2003). Monitoring involves judging how well information has been learned, and is especially important for self-regulated learning as monitoring affects subsequent control (or regulation) of the learning process. Control or regulation of the learning process can involve choices about which items need to be studied next or practiced further and how much time one should spend on them. For instance, research has shown that people tend to study longer on those items which they think they have not learned well (Metcalfe, 2009), and more accurate monitoring judgments have been found to lead to more accurate restudy choices and better final test performance (Thiede et al., 2003).

Judgments of learning (JOLs) are probably the most widely used monitoring judgments. JOLs require participants to either predict their memory for items on a future test (e.g., Nelson & Dunlosky, 1991) or to rate their comprehension of items (e.g., Thiede et al., 2003) during or after the learning phase and prior to taking that test. The difference between these two types of JOLs is usually related to the type of materials that are used in a study. That is, to monitor more complex materials such as expository text, a student should not only monitor memory but also whether he or she understood the text (i.e., comprehension).

In typical studies on monitoring accuracy using JOLs (see e.g., Anderson & Thiede, 2008; Dunlosky & Lipko, 2007; Koriat, Ackerman, Lockl, & Schneider, 2009a; Koriat, Ackerman, Lockl, & Schneider, 2009b; Metcalfe & Finn, 2008; Nelson & Dunlosky, 1991; Thiede et al., 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005), participants study items such as word pairs or expository texts, and provide JOLs by either predicting their future recall of each of the word pairs (e.g., Nelson & Dunlosky, 1991) or rating their comprehension of each of the texts (e.g., Maki, 1998b; Thiede et al., 2003). The relative accuracy of a judgment is then established by computing a Goodman–Kruskal gamma correlation between judgments and test performance, which can vary between -1 and $+1$; a gamma close to $+1$ would mean that criterion test performance on items that received higher recall/comprehension judgments was indeed better than performance on items that received lower judgments (Nelson, 1984). Relative accuracy measured by the gamma correlation indicates whether students can discriminate among items (i.e., whether items that get a higher JOL are indeed performed better on a test than items getting a lower JOL). Next to relative accuracy, monitoring accuracy can also be determined using absolute measures (Mengelkamp & Bannert, 2010; Schraw, 2009), in which the judgment for an item is compared with the performance on that item.

Given the important role that accurate monitoring was considered (and later established; e.g., Thiede et al., 2003) to play for effective self-regulation, it was problematic that early studies on word pairs and text often found monitoring accuracy to be quite low (e.g., Glenberg, Sanocki, Epstein, & Morris, 1987; Maki, 1998a; Nelson, Gerler, & Narens, 1984; Vesonder & Voss, 1985), and consequently, much subsequent research has focused on finding ways to improve monitoring accuracy.

1.2. Improving monitoring accuracy

In a well-known study with word pairs, Nelson and Dunlosky (1991) established that when asking about JOLs not immediately after studying a word-pair but after studying all word pairs, relative accuracy was higher. Nelson and Dunlosky called this the delayed-JOL effect, which they explained based on memory systems involved in making JOLs. Immediate JOLs might be less accurate because they are based on retrieval from both short-term memory (STM) and long-term memory (LTM), whereas delayed JOLs can only be based on LTM because STM traces of the item are no longer available. Schneider, Visé, Lockl, and Nelson (2000) found the delayed-JOL effect in primary school children from kindergarten, second grade, and fourth grade when learning word-picture pairs. Similar results were found by Koriat et al. (2009a, 2009b) with second and fourth grade primary school children learning word pairs. In their meta-analytic review, Rhodes and Tauber (2011) showed the robustness of this delayed-JOL effect on relative accuracy with paired associates, category exemplars, sentences and single words. Effect sizes for prospective memory items and information from videos were smaller. Also, the delayed-JOL effect was found to be less robust for children compared to other age groups.

Interestingly, however, early studies using texts suggested that the delayed-JOL effect did not apply to materials that were more complex than word pairs (Maki, 1998a). Maki (1998b) investigated text JOL accuracy under four conditions: (1) providing immediate JOLs and taking an immediate test after each text, (2) providing immediate JOLs but taking delayed tests after all texts were read and judged, (3) providing delayed JOLs and taking tests directly following the JOL about each text and (4) providing delayed JOLs and taking delayed tests after all JOLs were provided. The second condition is comparable to the immediate JOL and the fourth is comparable to the delayed JOL condition in the study by Nelson and Dunlosky (1991); but, in contrast to their study with word pairs, Maki's (1998b) data showed no difference in accuracy between those conditions (see Thiede et al., 2009, review for more studies that failed to find the delayed JOL effect with texts; e.g., Baker & Dunlosky, 2006; Dunlosky, Rawson, & Middleton, 2005).

However, as noted by Thiede et al. (2003), studies on monitoring accuracy with relatively simple tasks such as word pairs are different from studies on monitoring accuracy with more complex materials like texts. Task complexity can be defined in terms of element interactivity: the higher the number of interacting information elements that a learner has to relate and keep active in working memory when performing a task, the higher the complexity of that task and the higher the cognitive load it imposes (Sweller, Van Merriënboer, & Paas, 1998; Sweller, 2010). While learning word lists or word pairs requires memorization of isolated elements, learning texts requires building a mental representation consisting of multiple interacting elements. When providing a JOL about a text, then, learners have to judge the quality of their mental representation of the text, which differs markedly from JOLs about word pairs, which require learners to judge their ability to retrieve the learned information literally from memory. While simply delaying a judgment may provide a better cue for predicting memory of word pairs, it may not be sufficient for predicting the quality of the mental representation of a text.

Indeed, subsequent studies have shown that when participants are provided with instructions that focus their attention on the right cues (i.e., cues regarding the quality of their mental representation of the text) prior to making a comprehension judgment, their monitoring accuracy was enhanced. For example, generating keywords (Thiede et al., 2005) or making a summary (Anderson & Thiede, 2008) at a delay (i.e., after studying several texts) improved the relative accuracy of subsequent JOLs (Thiede et al., 2009). Similarly, immediate instructional strategies (i.e., after each text) such as rereading or self-explaining the text (Griffin, Wiley, & Thiede, 2008) or making a concept map of the text (Thiede, Griffin, Wiley, & Anderson, 2010) enhanced relative accuracy of immediate JOLs. What these different instructional strategies have in common, is that they provide learners with better diagnostic cues to assess their understanding or predict their test performance, by focusing their attention on their situation model (i.e., mental representation) of the text (Thiede et al., 2009). Because the situation model is the result of learners' understanding of the text and influences their test performance (Kintsch, 1998), JOLs should be based on cues from the situation model in order to be accurate (Rawson, Dunlosky, & Thiede, 2000; Wiley, Griffin, & Thiede, 2005).

Only a few studies investigated the accuracy of monitoring judgments (e.g., JOLs) in classroom settings. Several studies showed that children's monitoring judgments and regulation skills improved during school years (for a review see, Schneider, 2008). Primary school children were shown to be able to monitor whether their answers were correct or incorrect and regulate their learning accordingly, when