Learning and Individual Differences 41 (2015) 62–72

Contents lists available at ScienceDirect

Learning and Individual Differences

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

Adding judgments of understanding to the metacognitive toolbox☆

Celeste Pilegard *, Richard E. Mayer

University of California, Santa Barbara, United States

ARTICLE INFO

Article history:
Received 21 March 2014
Received in revised form 9 February 2015
Accepted 7 July 2015

Keywords:
Metacognition
Metacomprehension
Transfer
Multimedia learning

ABSTRACT

Two experiments investigated how the framing of metacognitive judgment prompts affects metacognitive accuracy. In Experiment 1, college students viewed a multimedia science lesson and were asked to make either judgments of learning (JOLs) or judgments of understanding (JOUs). The results indicated large correlations of JOUs with retention and transfer, and medium correlations of JOLs with retention and transfer. In Experiment 2, college students received the same lesson along with metacognitive prompts framed in terms of one's amount of knowledge ("how much"), one's confidence in knowledge ("how confident"), one's ability to answer questions ("how many"), or one's perceived difficulty in learning ("how difficult"). The former three judgments significantly predicted retention and transfer performance, but the judgment of difficulty did not significantly predict transfer. These results show the benefits of including judgments of understanding and transfer tests in studies that examine metacomprehension, and the importance of choosing appropriate wording for judgment prompts.

1. Introduction

Imagine you are an instructor designing a multimedia lesson. During the lesson, you want your students to gauge how well they are understanding the presented material. What question should you ask in order to prompt an accurate judgment? This is the question addressed in the present study.

The ability to accurately judge one's own learning of educational material is an important skill, especially because such judgments can influence study decisions (Son & Metcalfe, 2000; Thiede, Anderson, & Therriault, 2003; Winne & Hadwin, 1998). It is critical, then, to prompt a learning judgment with a question whose answer correlates most highly with actual learning outcomes. The first experiment aims to systematically assess two factors that could affect the magnitude of this correlation: the framing of the judgment prompt (i.e., as how well you remember versus how well you understand) and the type of learning outcome measured (i.e., retention versus transfer). The second experiment investigates whether metacognitive accuracy is affected by different framings of prompts, such as prompts based on magnitude of knowledge (how well do you know?), certainty of knowledge (how confident are you that you know?), ability to answer test questions (how many questions will you answer correctly?), or difficulty of acquiring knowledge (how difficult was learning?). These factors are evaluated using a computer-based multimedia lesson involving a difficult science topic.

1.1. Literature review

A classic approach to understand the relationship between what learners know and what they think they know is the Judgment of Learning (JOL) paradigm (e.g., Arbuckle & Cuddy, 1969). In this paradigm, participants study a list of arbitrarily paired associates. They then make a judgment of the likelihood that, given the first word, they will remember the second word on a test. In this standard paradigm participants are generally inaccurate at gauging their knowledge. Adjustments to the paradigm such as adding a delay between learning and making a judgment of learning have resulted in increased accuracy (Nelson & Dunlosky, 1991) though the nature of this increase is disputed (Kimball & Metcalfe, 2003). Extensive research has used this JOL framework to advance the understanding of metacognitive accuracy for rote learning (Dunlosky & Metcalfe, 2009).

While the paired associate paradigm has educational analogs in areas such as foreign vocabulary learning, it does not allow researchers to gauge comprehension of potentially meaningful material. The distinction between the processes underlying retention and the processes underlying comprehension has been espoused by numerous theories. In the framework established by Kintsch and colleagues, the processing of discourse occurs at strategic levels in a hierarchy of complexity (Kintsch, 1998; Van Dijk & Kintsch, 1983). The least complexity is contained in the *surface form*, which involves simply representing the words exactly as they are presented and has a very high rate of decay. The next level is the *textbase*, which is somewhat less literal and involves minor semantic inferences. The most semantically complex level of comprehension is

☆ Author Note: This research was supported by grants from the Office of Naval Research under Grant No. N000141110225 and the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1144085. Carley Ward assisted on the literature review for this project.

* Corresponding author at: Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106-9660.
E-mail address: pilegard@psych.ucsb.edu (C. Pilegard).

http://dx.doi.org/10.1016/j.lindif.2015.07.002
1041-6080/

the *situation model*, which involves mentally representing the order and layout of a situation and requires integrating the presented information with prior knowledge. In this framework, creating a situation model is analogous to learning for *understanding* and creating a textbase is analogous to learning for *retention*. In this sense, comprehension goes beyond presented information to allow the learner to make inferences and build models. Someone who does not construct a proper situation model cannot be said to comprehend the presented information. Paired-associate tests, vocabulary-definition tests, and fact reproduction tests can only gauge rote retention, and therefore lack the means to gauge comprehension. Similarly, word lists, vocabulary sheets, and factoid-based materials lack any underlying structure to make comprehension possible.

The distinction between learning for retention (or rote learning) and learning for understanding (or meaningful learning) can be conceptualized in terms of the underlying processing. According to the cognitive theory of multimedia learning (Mayer, 2005, 2009), *meaningful learning* requires actively selecting (i.e., paying attention to the important information), organizing (i.e., creating a comprehensible structure), and integrating (i.e., making connections with prior knowledge) information. In contrast, rote learning involves only the cognitive process of selecting, without organizing or integrating. The processes in meaningful learning can also be characterized as *generative processing*, which requires the active construction of knowledge, whereas the processes in rote learning can be characterized as *reproductive processing*, which involves processing information without elaboration (Mayer, 2005, 2009; Wittrock, 1990; Wittrock & Carter, 1975). This distinction between learning by rote and learning by understanding has its roots in the Gestalt work of Katona (1940) and Wertheimer (1945).

In order to assess retention and understanding, it is necessary to develop tests of knowledge that require either remembering the presented information or the ability to apply the information to new situations. *Retention tests* require only the remembering of information. These tests involve questions that can be answered by reproducing or recognizing the information presented in the lesson. *Transfer tests* require using the presented information in novel ways that were not directly addressed in the lesson (Mayer & Wittrock, 1996, 2006). Well-designed transfer questions require understanding of the material and cannot be correctly answered with rote memory alone.

The field of *metacomprehension* seeks to investigate the accuracy with which people can gauge their comprehension, beyond gauging their memory (Dunlosky & Lipko, 2007). The distinction between rote memory and understanding is an important distinction to make from an educational standpoint, as most teachers hope to instill their students with meaningful understanding of class material rather than solely rote retention (Mayer, 2002, 2009). The standard paradigm for metacomprehension studies was first developed by Glenberg and Epstein (1985). This paradigm is sometimes referred to as *calibration of comprehension* (Lin & Zabrucky, 1998). Participants typically read a number of expository texts and make some sort of comprehension judgment for each one. This is followed by a test of the material that includes questions corresponding to each text. Accuracy is typically calculated with a within-subject correlation between learning judgments and test performance. The typical metacomprehension study uses text passages for study material. In general, other types of educational materials (such as multimedia lessons or computer-based materials) have not been included (but for exceptions see Ackerman & Goldsmith, 2011; Serra & Dunlosky, 2010), despite calls to replicate and extend research findings in new contexts (Shavelson & Towne, 2002). In the present study, we investigate metacognitive accuracy in the context of a narrated computer-based multimedia science lesson, which describes a causal system. We focus on academic content that involves a cause-and-effect system because it allows for learners to develop a deeper understanding than a collection of facts, which is commonly used in metacognition research.

Metacomprehension judgments are typically inaccurate (Dunlosky & Lipko, 2007; Lin & Zabrucky, 1998). Many attempts have been made to increase the accuracy of these judgments, especially because accurate judgments have been shown to lead to more effective study decisions (Son & Metcalfe, 2000; Thiede et al., 2003). Some studies have found moderate success in improving metacomprehension accuracy through study techniques such as summarizing or generating keywords for the material (Thiede & Anderson, 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005). In general, however, improvements to metacomprehension accuracy are elusive. Many techniques that improve JOL accuracy (e.g., adding a delay) either don't improve metacomprehension accuracy or result in inconsistent effects, (Maki, 1998; Maki & Berry, 1984).

Wiley, Griffin, and Theide (2005) proposed that the elusiveness of accurate metacomprehension judgments could be due to problems with the experimental materials themselves. Their two major concerns are that assessing metacomprehension is impossible if: (1) metacomprehension studies are performed with texts that only allow for memory rather than comprehension (such as simple factoid-based texts with few structural relations), and (2) tests of learning only require memory for the texts rather than comprehension. It is possible that some studies that have sought to test metacomprehension have actually tested metamemory because of the types of texts used. To address the first problem, they recommend the use of texts with an underlying causal model so that subjects can build and evaluate a situation model as a part of their comprehension judgment. To address the second problem, they recommend using test questions that require inferences. We adhere to these recommendations in the present study by using a science text with a causal chain of events and a test containing transfer questions that require using the presented material to solve novel problems.

Another inconsistency in the literature on metacomprehension is the focus of the judgment prompt. Judgment prompts differ in whether they explicitly ask participants to consider their comprehension or their understanding. A judgment prompt asking a participant to gauge how many test questions they will answer correctly, for example, could specifically mention a comprehension test: “What percentage of comprehension questions do you expect to answer correctly?” (Ackerman & Goldsmith, 2011) or not mention the test type: “Please indicate how many of the five test questions you think you will answer correctly on the text” (Redford, Thiede, Wiley, & Griffin, 2012).

Some calibration studies have included different prompt types and learning outcome measurements. Rawson, Dunlosky, and McDonald (2002) compared learners' predictions about test performance against comprehension judgments. They found that comprehension judgments and performance predictions behave differently: for example, comprehension judgments were higher in magnitude than performance predictions, and performance predictions were affected by expected test delay while comprehension judgments were not. These findings suggest that different types of prompts may tap different cognitive processes. In a study on the effect of reading from a computer on metacognitive regulation, Ackerman and Goldsmith (2011) used a within-subjects design to elicit comprehension and memory judgments, and measured learning with multiple-choice questions that required either inferences or memory for details. Unfortunately the two ratings and two types of performance were averaged in their analysis, and any differences between them were not reported. Lin, Moore, and Zabrucky (2001) elicited judgments of understanding, confidence, easiness, and interestingness for texts, and used four true–false inference questions to measure learning. Their results showed that the four judgments were highly correlated with one another, and none of the judgments stood out as producing more accurate ratings than any other.

Another inconsistency is the general framing of the judgment prompt. Participants may be asked to indicate the degree to which they understand the material (e.g., Serra & Dunlosky, 2010; Walczyk & Hall, 1989), their confidence in their ability to answer test questions (e.g., Bouffard-Bouchard, 1994; Glenberg & Epstein, 1985), how many test questions they will answer correctly (e.g., Ackerman & Goldsmith, 2011; Redford et al., 2012), or how easy they found the text (e.g., Maki, Foley, Kajer, Thompson, & Willert, 1990; Maki & Serra,