Learning and Instruction 34 (2014) 58–73

Contents lists available at ScienceDirect

Learning and Instruction

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

# Do illustrations help or harm metacomprehension accuracy?

Allison J. Jaeger*, Jennifer Wiley [1]

*University of Illinois at Chicago, 1007 West Harrison Street M/C 285, Chicago, IL 60607, USA*

ARTICLE INFO

*Article history:*
Received 31 December 2013
Received in revised form
29 July 2014
Accepted 10 August 2014
Available online 28 August 2014

*Keywords:*
Metacomprehension
Self-explanation
Multimedia

ABSTRACT

The present research examined the effect of illustrations on readers' metacomprehension accuracy for expository science text. In two experiments, students read non-illustrated texts, or the same texts illustrated with either conceptual or decorative images; were asked to judge how well they understood each text; and then took tests for each topic. Metacomprehension accuracy was computed as the intra-individual correlation between judgments and inference test performance. Results from both studies showed that the presence of decorative images can lead to poor metacomprehension accuracy. In the second study, an analysis of the cues that students reported using to make their judgments revealed that students who used comprehension-relevant cues showed more accurate metacomprehension. A self-explanation instruction did not alter either comprehension-relevant cue use or metacomprehension accuracy, although some advantages were seen when readers were prompted to self-explain from texts illustrated with conceptual images. These results suggest that students may need more explicit instruction or support to promote the use of valid cues when engaging in comprehension monitoring with illustrated text, and that seductive information such as decorative images may undermine comprehension monitoring.



## 1. Introduction

Accurate monitoring of comprehension (i.e. being able to differentiate between topics that have been understood well after reading from those which have not been understood well) is critical to successful self-regulated learning (Griffin, Wiley, & Salas, 2013; Thiede, Anderson, & Therriault, 2003). On any given night of homework, a student may need to read text passages about early civilizations such as the Aztecs and Incas to prepare for one test, as well as passages about photosynthesis and ecosystems to prepare for another. Effective self-regulation is especially important in these situations because it is by monitoring one's own progress while learning that decisions are made about what material needs to be restudied. If students are unable to accurately differentiate between well-learned material and less-learned material, they may waste time returning to material that is already well understood. Given the limited amount of time available for study, this may mean they will fail to restudy material that is not well understood. Despite the importance of accurate monitoring for effective self-regulated learning, students are generally poor at assessing their understanding of text passages, with typical correlations between predicted test performance and actual test performance being around .27 (Dunlosky & Lipko, 2007; Maki, 1998a; Thiede, Griffin, Wiley, & Redford, 2009). With the increasing popularity and ease of creating multimedia presentations for information in this digital age and given the widespread use of multimedia materials in educational settings, it is an important question how multimedia adjuncts may alter the metacomprehension process. Therefore, the main purpose behind this set of studies is to explore how adding illustrations to expository science texts may either improve or harm comprehension monitoring accuracy.

### *1.1. What is metacomprehension accuracy?*

Comprehension monitoring accuracy or *metacomprehension accuracy* refers to the ability of an individual to predict how well one will do on a set of comprehension tests after reading a set of texts. Several measures of metacomprehension compare metacognitive judgments with actual performance, but each one does so in a slightly different manner. These measures include absolute accuracy, confidence bias, and relative accuracy (Maki, 1998a). Absolute accuracy is computed as the mean absolute deviation between judged and actual performance. This measure is sometimes referred to as calibration because it gives an idea of how far off a person's judgments are from actual performance. Confidence

* Corresponding author. Tel.: +1 312 339 7677; fax: +1 312 413 4122.
*E-mail addresses:* ajaege1@uic.edu (A.J. Jaeger), jwiley@uic.edu (J. Wiley).
[1] Tel.: +1 312 355 2501; fax: +1 312 413 4122.

http://dx.doi.org/10.1016/j.learninstruc.2014.08.002
0959-4752/

bias is a similar measure but actually concerns the direction of people's misjudgments and is sometimes referred to as over-/under confidence. This measure is computed as the signed difference between mean judgments and mean performance. Finally, relative accuracy, which is sometimes referred to as discrimination accuracy or monitoring resolution, refers to a participant's accuracy in predicting performance on one text relative to other texts (Glenberg & Epstein, 1985; Maki & Berry, 1984). As recommended by Nelson (1984), relative monitoring accuracy is computed as an intra-individual correlation between readers' judgments of learning for each text relative to the other texts, and their actual performance on each test relative to other tests. Correlations can range from −1 to +1, with correlations near 0 or below representing chance to poor accuracy. Correlations near +1 would indicate very good discrimination between texts one has understood well from those one has not. To make this concrete, imagine again a student has 4 reading assignments on a given night on the topics of Aztec Civilizations, Incan Civilizations, Photosynthesis, and Ecosystems. After having engaged in a first pass of studying, a student could be asked to rate their understanding of the 4 texts. Let's say they indicate having understood the Aztec text the best, then the Inca text, then Photosynthesis and then Ecosystems. The student would then be given test questions on each of these topics. If the test scores are aligned with the predictions (i.e. 90%, 80%, 70% and 60%) the student would be said to have perfect relative accuracy.

Although all three measures of metacomprehension accuracy are similar in that they pertain to how well a person's judgments are related to their target performance, absolute accuracy and confidence bias are statistically independent from relative accuracy (Dunlosky & Thiede, 2013). For example, a student can have good absolute accuracy or confidence bias, but poor relative accuracy. Further, absolute accuracy and confidence bias can be influenced by factors that do not affect relative accuracy measures. Specifically, absolute accuracy and confidence bias are dependent upon mean performance levels (Nelson, 1984). This can be problematic because it can in turn allow for non-metacognitive factors to influence the accuracy scores obtained, for example by things such as text or test difficulty and amount of prior knowledge (Griffin, Jee, & Wiley, 2009). Because relative accuracy is less affected by non-metacognitive factors, it is this measure that has been most commonly used in studies of metacomprehension accuracy while learning from text following the tradition established in seminal work by Glenberg and Maki (Dunlosky & Lipko, 2007; Glenberg & Epstein, 1987; Maki, 1998b; Maki & Serra, 1992; Nelson & Dunlosky, 1991; Thiede et al., 2009) and is also the measure that will be employed in the current studies.

### 1.2. Basic model of metacomprehension accuracy

As mentioned earlier, relative metacomprehension accuracy is partly determined by the judgments that are made by a reader. Koriat (1997) proposed the cue-utilization account to explain the accuracy of judgments-of-learning (JOLs) as a function of the cues that are used as the basis for judgments. This account posits that people have a variety of cues that they can use to predict their own test performance, and that the accuracy of these predictions hinges upon whether the chosen cues are consistent with the factors that will actually affect performance on the tests.

There is an extensive literature looking at JOLs and memory test performance for learning paired-associates such as words and their definitions and foreign language vocabulary (Metcalfe, 2002; Metcalfe & Kornell, 2003; Nelson & Dunlosky, 1991). One of the most robust findings from this literature indicates that delaying judgments serves to substantially increase the relative accuracy of JOLs compared to JOLs solicited immediately after study (see Rhodes & Tauber, 2011 for a review). Work in this area has also consistently shown that JOLs are higher for related items than unrelated items (Dunlosky & Matvey, 2001). Further, this literature has also shown support for the idea that people can make strategic study decisions based on their metacognitions (Metcalfe, 2009), a finding that demonstrates the importance of monitoring.

Although the cue-utilization account was originally formulated to explain predictions of performance in metamemory paradigms where participants are predicting their ability to recall a learned item from memory, it has also been useful in understanding the mechanisms that may be underlying meta*comprehension* accuracy, where participants are predicting whether they have learned the information that has been presented in a text. In studies that have explored metacomprehension accuracy, it has been argued the nature of "learning" that needs to be judged in the case of learning from text differs fundamentally from the previous work that used JOLs for paired-associates learning tasks. Text researchers, building on the work of Kintsch and Van Dijk (1978), have pointed out that "learning" from a text requires both memory for the text and understanding the meaning of the text, which occurs via the construction of a situation-model level representation (Rawson, Dunlosky, & Thiede, 2000; Wiley, Griffin, & Thiede, 2005). Thus, when asked to predict one's learning of a text, the task becomes more complicated than when one is asked to predict their memory performance. When asked to make JOLs when learning from text, readers have access to many cues that could affect how these judgments are made. As might be expected, one main cue they tend to use is their memory for the text (Rawson, Dunlosky, & McDonald, 2002; Thiede, Griffin, Wiley, & Anderson, 2010). However, in addition, readers also tend to rely on heuristic cues such as their interest in the topic, their prior knowledge or familiarity with the topic, or feelings of fluency while reading (Griffin et al., 2009; Rawson et al., 2000; Thiede et al., 2010) when making predictions. While these types of cues may be very salient to a reader, they are not directly related to the process of creating a mental model of the text and therefore are likely to be less valid predictors of performance on comprehension tests (Dunlosky, Rawson, & Middleton, 2005; Griffin et al., 2013; Wiley et al., 2005). The use of these cues may be responsible for the generally poor levels of metacomprehension accuracy that have been observed, around .27 (Thiede et al., 2010).

Other cues, referred to as representation-based cues (Thiede et al., 2010), develop from the process of attempting to create a mental model or situation-model-level representation of that text. These cues could include whether the person feels they could summarize the process described by the text or explain it to someone else. Although these cues are better predictors of comprehension, they tend to be used less often by students when making comprehension judgments (Thiede et al., 2010). Despite the general tendency for readers to make inaccurate judgments about comprehension, several studies have shown notable improvements in metacomprehension accuracy by putting readers in contexts designed to invoke the use of situation-model-based cues (Thiede et al., 2009). For example, readers have been shown to be more accurate when they generate keywords or summaries after a delay (Thiede & Anderson, 2003; Thiede et al., 2003). The mechanism that is suggested to underlie this phenomenon is that as time passes, surface cues decay and become less accessible, while the situation model is more robust to forgetting (Kintsch, Welsch, Schmalhofer, & Zimny, 1990). So, when keywords or summaries are generated after a delay, it helps readers to access more valid situation-model-based cues (Thiede, Dunlosky, Griffin, & Wiley, 2005). Similarly, having readers create concept maps or self-explain as part of reading has been shown to increase