



To What Extent Do Situation-Model-Approach Interventions Improve Relative Metacomprehension Accuracy? Meta-Analytic Insights

Anja Prinz¹ · Stefanie Golke¹ · Jörg Wittwer¹

Published online: 14 October 2020
© The Author(s) 2020

Abstract

This meta-analysis investigated the extent to which relative metacomprehension accuracy can be increased by interventions that aim to support learners' use of situation-model cues as a basis for judging their text comprehension. These interventions were delayed-summary writing, delayed-keywords listing, delayed-diagram completion, self-explaining, concept mapping, rereading, and setting a comprehension-test expectancy. First, the general effectiveness of situation-model-approach interventions was examined. The results revealed that, across 28 effect sizes (comprising a total of 2,236 participants), situation-model-approach interventions exerted a medium positive effect ($g = 0.46$) on relative metacomprehension accuracy. Second, the interventions were examined individually. The results showed that, with the exception of self-explaining, each intervention had a significant positive effect on relative metacomprehension accuracy. Yet, there was a tendency for setting a comprehension-test expectancy to be particularly effective. A further meta-analysis on comprehension in the selected studies revealed that, overall, the situation-model-approach interventions were also beneficial for directly improving comprehension, albeit the effect was small. Taken together, the findings demonstrate the utility of situation-model-approach interventions for supporting self-regulated learning from texts.

Keywords Interventions · Meta-analysis · Metacomprehension accuracy · Relative accuracy · Situation-model approach

✉ Anja Prinz
anja.prinz@ezw.uni-freiburg.de

¹ Department of Educational Science, University of Freiburg, Rempartstrasse 11, 79098 Freiburg, Germany

Introduction

Much of the learning students engage in is based on reading text material. For example, students complete reading assignments, review class notes, and read books and articles for writing a term paper. When doing so, they must monitor and judge their comprehension of the material. Students who more accurately judge which material they have comprehended well and which material they have comprehended less well—that is, students who exhibit greater *relative metacomprehension accuracy*—can better regulate their learning, for instance, by selectively restudying. This adaptive regulation in turn results in an improved comprehension (e.g., Rawson et al. 2011; Shiu and Chen 2013; Thiede et al. 2003, 2012). However, learners typically reveal poor relative metacomprehension accuracy (see, e.g., Prinz et al. 2020; Thiede et al. 2009). One major reason for inaccurate metacomprehension is that learners base their judgments on inappropriate cues (e.g., quantity of recallable details; Thiede et al. 2010). Hence, interventions to support learners to use more appropriate cues (e.g., reproducibility of key ideas) have been developed, such as delayed-summary writing, delayed-keywords listing, delayed-diagram completion, self-explaining, concept mapping, rereading, and setting the right test expectancy (see, e.g., Griffin et al. 2013, 2019a; Wiley et al. 2016b). Yet, to date, the extent to which these interventions are effective to improve relative metacomprehension accuracy is unclear. Therefore, we conducted a meta-analysis to uncover and compare the effectiveness of the interventions. The present meta-analysis extends previous narrative reviews in important ways. First, clear sets of rules were predefined to search and select studies for inclusion, and sophisticated statistical methods (e.g., weighting effect sizes by their precision) were used to aggregate the selected studies. This led to increased statistical power and a reliable estimate of the average effect size (see, e.g., Higgins and Green 2011). In addition, the meta-analysis indicates to what extent the effect is consistent across the studies (see, e.g., Borenstein et al. 2009). Second, this meta-analysis has the potential to contribute to greater evidence-based practice. There has been the call for education to place increased emphasis on empirical evidence as basis for making decisions about instructional practices. A central requirement for such evidence-based practice are methodologically sound meta-analyses that synthesize findings to reveal effective methods (see, e.g., Slavin 2008).

Difficulties with and Importance of Relative Metacomprehension Accuracy

The earliest studies that used and established the typical relative metacomprehension-accuracy research paradigm were conducted by Maki and Berry (1984) and Glenberg and Epstein (1985). In this paradigm, participants read several texts, predict their comprehension of each text, and complete a test on each text. Relative metacomprehension accuracy is then operationalized as the intraindividual correlation between a participant's predictions and actual test performance scores across the set of texts. Typically, Gamma or Pearson correlation coefficients are computed. These coefficients range from -1.00 to $+1.00$, with stronger positive correlations indicating greater accuracy (see, e.g., also Griffin et al. 2019a).

Research has demonstrated that learners are typically poor at accurately discriminating between more and less well-understood texts. A recent meta-analysis showed that relative metacomprehension accuracy is on average as low as $+0.24$ (i.e., without any intervention or support; Prinz et al. 2020). This is in line with previous narrative reviews suggesting that the average level of relative metacomprehension accuracy is between $+0.20$ and $+0.30$ (Dunlosky and Lipko 2007; Lin and Zabricky 1998; Maki 1998; Thiede et al. 2009; Weaver et al. 1995).

However, it is important that learners accurately judge their comprehension of texts because it enables them to effectively regulate their studying by devoting time and resources to where they are needed (e.g., De Bruin et al. 2011; Schleinschok et al. 2017; Thiede et al. 2003, 2012). Consequently, to support text-based learning, it is crucial to discover ways to improve relative metacomprehension accuracy.

The Role of Cue Utilization for Relative Metacomprehension Accuracy

The *cue-utilization framework* provides an explanation for why relative metacomprehension accuracy is commonly poor (Griffin et al. 2009; cf. Koriat 1997; for an overview of constraints on relative metacomprehension accuracy, see Thiede et al. 2009). This framework supposes that learners do not have direct access to their cognitive states but have to infer their level of comprehension based on cues. In doing so, they can use a variety of cues, and judgment accuracy depends on how strongly the cues used are tied to the mental text representation that determines performance on the respective comprehension test. Heuristic cues, such as domain familiarity (e.g., Glenberg et al. 1987) or topic interest (e.g., Lin et al. 1996), are available whether or not a text has been read and are therefore insensitive to idiosyncrasies of a specific mental text representation. Hence, they typically yield inaccurate judgments. Representation-based cues, such as ease of processing (e.g., Maki et al. 1990) or accessibility of textual information (e.g., Baker and Dunlosky 2006), become available only during or after reading and are therefore more closely related to a particular mental text representation. Thus, they generally yield more accurate judgments.

Nonetheless, not all representation-based cues are equally valid indicators of comprehension. Their validity depends on the correspondence between the level of mental text representation the cues are related to and the level of mental text representation tapped by the comprehension test. According to the *construction-integration model of text comprehension* (Kintsch 1998), a learner concurrently constructs a mental representation of a text on multiple levels: The surface level is constructed as the words are encoded. The textbase level is constructed as the words are parsed into propositions and as links between the propositions are formed. The situation-model level is constructed as different pieces of textual information are connected with each other and with the learner's prior knowledge. The situation model represents a learner's deep understanding and what is typically referred to as *text comprehension*. Thus, a comprehension test should assess a learner's situation model by requiring inferences or application of the textual information rather than a learner's surface or textbase representations by demanding the recall of specific textual information (which is often referred to as metamemory instead of metacomprehension, see, e.g., Griffin et al. 2013, 2019a; for a detailed discussion on designing texts and tests for assessing the situation model, see Wiley et al. 2005). When a comprehension test actually assesses the situation model, metacomprehension will be less accurate if learners use cues tied to their surface or textbase representations but more accurate if learners use cues that tap their situation model. This perspective is called the *situation-model approach to metacomprehension* (Griffin et al. 2019a; Wiley et al. 2016b). However, Thiede et al. (2010, Experiment 1; see also Jaeger and Wiley, 2014, Experiment 2) gathered self-reports of cue use and found that, rather than using situation-model cues, learners tend to default to heuristic cues or representation-based cues related to their memory of specific textual information, leading to less accurate judgments.

Potential explanations for why learners often focus on heuristic and memory-based cues are provided by the *effort monitoring and regulation framework* (De Bruin et al. 2020). For one,

this framework suggests that monitoring and regulation can be impaired through unnecessary cognitive load imposed by the learning task or the inadequate distribution of load between the task and metacognitive processes. Such resource constraints might force learners to draw on more easily available heuristic and memory-based cues (cf. Griffin et al. 2009). In addition, the framework suggests that learners sometimes misinterpret cues, in particular, their invested mental effort. Research has indicated that, although high mental effort does not necessarily indicate that learning is ineffective, learners tend to judge their learning as low when they experience high mental effort (e.g., Baars et al. 2014, 2018, Experiment 2; Dunlosky et al. 2006; Koriat et al. 2009; Schleinschok et al. 2017; see also Baars et al. 2020; Carpenter et al. 2020). Hence, learners are confronted with the difficulty that their invested mental effort is a salient but not necessarily predictive cue of their actual learning and particularly their deeper comprehension.

Situation-Model-Approach Interventions to Improve Relative Metacomprehension Accuracy

Following the situation-model approach, interventions have been developed to increase relative metacomprehension accuracy by encouraging learners to use cues that are related to their situation model. More precisely, these situation-model-approach interventions are designed to help learners *generate*, *attend to*, or *select* situation-model cues (see, e.g., Griffin et al. 2013, 2019a; Wiley et al. 2016b).

Cue-Generation Interventions

Some situation-model-approach interventions support the generation of situation-model cues. They do so either by requiring *encoding* on the situation-model level, as is the case for self-explaining and concept mapping, or by requiring *retrieval* of the situation model, as is the case for delayed-generation tasks (Thiede et al. 2019).

Delayed Generation of Summaries, Keywords, and Diagrams Research has shown that delayed-generation tasks, such as writing a summary of each text, listing keywords that capture the essence of each text, and completing a diagram for each text, can lead to enhanced relative metacomprehension accuracy. More precisely, when these tasks were completed after a short delay, that is, after all texts had been read, relative metacomprehension accuracy was higher than when they were completed immediately after reading each text or when no task was completed (delayed-summary writing: Anderson and Thiede 2008; Engelen et al. 2018, Experiment 1; Thiede and Anderson 2003; Thiede et al. 2010, Experiment 1; delayed-keywords listing: De Bruin et al. 2011; Engelen et al. 2018, Experiment 1; Shiu and Chen 2013; Thiede et al. 2003, 2005, 2012, Study 2, 2017; Waldeyer and Roelle, 2020; delayed-diagram completion: Van de Pol et al. 2019, Experiment 1; Van Loon et al. 2014; see also Van de Pol et al. 2020). It is important to note that neither delayed-nongenerative tasks (i.e., thinking about a text or reading keywords) nor simply delaying predictions affected relative metacomprehension accuracy (Thiede et al. 2005). Consequently, both the *generation* and the *delay* between reading and the task play a crucial role in improving relative metacomprehension accuracy. First, a generation task encourages learners to access a representation of the text from memory prior to making a judgment. Specifically, writing a summary, listing keywords, or completing a diagram functions as a kind of self-test, providing

learners with cues such as how successfully they can retrieve textual information. Second, the surface and textbase representations of a text fade quite rapidly, whereas the situation model is more stable over time (Kintsch et al. 1990). Hence, when performing a generation task immediately after reading, learners have easy access to and may primarily use their surface and textbase representations. However, the cues obtained from this experience are of limited predictive validity for later comprehension-test performance, which is largely determined by the situation model. In contrast, performing a generation task after a delay allows for the surface and textbase representations to decay, forcing learners to access and use their situation model. The cues gained from this experience are more predictive of later comprehension-test performance and therefore improve relative metacomprehension accuracy (see, e.g., also Griffin et al. 2013, 2019a).

Some studies provided support for the proposed mechanisms underlying the effectiveness of delayed-generation tasks by investigating learners' cue use (Anderson and Thiede 2008; Engelen et al. 2018, Experiment 1; Thiede et al. 2010, Experiment 1; Thiede and Anderson 2003; Van de Pol et al. 2019, Experiment 1; Van Loon et al. 2014). For example, Anderson and Thiede (2008) conducted qualitative analyses of the generated summaries. The results indicated that, when writing summaries immediately, the learners frequently based their predictions on how well they felt they could reproduce details reported in a text. In contrast, when writing summaries after a delay, the learners often used their perception of how well they could describe gist ideas of a text to predict their comprehension. Thus, delayed-summary writing enhanced the learners' use of situation-model cues, producing a higher accuracy level (see also Thiede and Anderson 2003; Van de Pol et al. 2019, Experiment 1; Van Loon et al. 2014). Furthermore, Thiede et al. (2010, Experiment 1) had learners report what cues they used when making their predictions following delayed-, immediate-, and no-summary writing. Most learners reported relying on how much textual information they could recall. This cue supported accurate predictions only in the delayed-summary condition. This result indicates that delayed-generation tasks render representation-based cues more valid by shifting focus to the situation model. However, although some studies found that delayed-keywords listing is also effective to increase relative metacomprehension accuracy for primary- and secondary-school students (De Bruin et al. 2011, Experiments 1 and 2; Thiede et al. 2017), some studies did not find benefits for these younger learners (De Bruin et al. 2011, Experiment 2; Engelen et al. 2018, Experiment 1; Pao 2014). Similarly, for delayed-summary writing, one study did not find a benefit for primary-school students (Engelen et al. 2018, Experiment 1). This suggests that the favorable impact of delayed-generation tasks may not consistently occur. Presumably, the quality of the keywords (e.g., whether they refer to central ideas or unimportant facts), summaries (e.g., whether they cover all main ideas), or diagrams (e.g., whether relations are clear) as well as learners' cognitive capacity to generate them plays a role for the tasks' effectiveness (cf., e.g., De Bruin et al. 2011; Roebbers et al. 2007).

Self-Explaining Self-explaining is the activity of explaining to oneself the meaning of information presented in a text as well as how it fits together with other information and the overall theme (e.g., Chi 2000). Self-explaining during reading can increase relative metacomprehension accuracy (Griffin et al. 2008, Experiment 2, 2019b, Experiment 4; Wiley et al. 2016a¹; Wiley et al. 2008, Experiment 2). By considering the relevance of textual

¹ Note, however, that this study compared an intervention condition receiving a combined comprehension-test-expectancy and self-explanation intervention with a control condition not receiving any intervention. In addition, the learners had the opportunity to regulate their learning before taking the tests.

information and trying to make inferences during self-explaining, learners simultaneously construct and reflect on their situation model, generating situation-model cues. Hence, in contrast to delayed-generation tasks, which prompt learners to draw on an already constructed situation model (i.e., retrieval-based cue-generation intervention), self-explaining encourages learners to focus on their situation model during reading (i.e., encoding-based cue-generation intervention; Thiede et al. 2019). Relatedly, in contrast to delayed-generation tasks, which need to be performed after a time lag to increase access to situation-model cues, self-explaining directly involves the situation model, making a delay superfluous (see, e.g., also Griffin et al. 2019a; Wiley et al. 2016b). Griffin et al. (2008, Experiment 2) found a positive effect of self-explaining on relative metacomprehension accuracy above mere rereading and independent of individual differences in working-memory capacity and reading skill. This finding suggests that the effect of self-explaining is not tied to the alleviation of processing constraints that can impede monitoring but rather attributable to the generation of situation-model cues. However, in a study by Jaeger and Wiley (2014, Experiment 2), self-explaining failed to improve relative metacomprehension accuracy, indicating that the effect of this intervention may be of limited reliability.

Concept Mapping Concept mapping is the activity during which learners visually depict the connections among concepts presented in a text (e.g., Weinstein and Mayer 1986). Concept mapping during reading has been found to enhance relative metacomprehension accuracy (Redford et al. 2012, Experiment 2; Thiede et al. 2010, Experiment 2). Concept mapping may be effective for two reasons. First, the activity yields an external visual representation of a learner's understanding. Thus, when, for example, trying to comprehend later parts of a text, a learner can review earlier content on the concept map. This eases working-memory demands that can be devoted to monitoring. Second, similar to self-explaining, concept mapping promotes learners not only to build a situation model but also to reflect on it, generating situation-model cues (i.e., encoding-based cue-generation intervention; Thiede et al. 2019; see, e.g., also Redford et al. 2012; Wiley et al. 2016b). Supporting this assumption, Thiede et al. (2010, Experiment 2) found that learners used their perception of how many appropriate connections between concepts they could make in a concept map as a situation-model cue for predicting their comprehension, which led to an increased accuracy level. Moreover, Redford et al. (2012, Experiment 2) revealed that only learners who constructed concept maps but not learners who received completed concept maps achieved greater relative metacomprehension accuracy than learners in a no-intervention control group. This outcome provides evidence that the act of constructing concept maps is critical because it enables learners to generate relevant cues.

Cue-Attention Intervention

A further intervention that can promote relative metacomprehension accuracy is repeated reading, which targets at facilitating the attention to situation-model cues (Dunlosky and Rawson 2005; Griffin et al. 2008, Experiment 1²; Rawson et al. 2000). During a first reading, learners may often lack the cognitive resources to concurrently engage in comprehension and monitoring processes and therefore primarily rely on heuristic cues. In contrast, during a

² The same trend was seen in Experiment 2 but was not statistically significant.

second reading, many of the processes involved in comprehension do not have to be re-executed (e.g., Millis et al. 1998). This allows learners to put more resources into careful monitoring (see, e.g., also Griffin et al. 2019a). Accordingly, Griffin et al. (2008) found that the relative metacomprehension accuracy of learners with low working-memory capacity or low reading skill was limited after a single reading. However, after rereading, their accuracy level was increased such that they were as accurate as learners with high working-memory capacity or high reading skill. The authors argued that rereading provides a second chance at monitoring, which is especially beneficial for learners whose limited cognitive capabilities prevent them from thorough monitoring during the first pass (for alternative but empirically less supported explanations for the rereading effect, see also Dunlosky and Rawson 2005; Rawson et al. 2000). In addition, Rawson et al. (2000) ruled out the possibility that improved relative metacomprehension accuracy after rereading is an artifact of familiarity with all texts prior to providing predictions or of increased test reliability. Furthermore, as indicated previously, when rereading was combined with a self-explanation intervention, both were independently beneficial for enhancing relative metacomprehension accuracy. This outcome supports the theoretical distinction between the two mechanisms: Interventions like self-explaining enable the generation of situation-model cues, whereas rereading enhances the attention to the cues (Griffin et al. 2008, Experiment 2). It should be noted, however, that some studies did not find a benefit of rereading over reading once (Chiang et al. 2010; Margolin and Snyder 2018), and, in some studies, learners exhibited low relative metacomprehension accuracy despite rereading (Bugg and McDaniel 2012; Pao 2014; Redford et al. 2012, Experiment 1).

Cue-Selection Intervention

A common feature of the interventions described so far is that they prompt learners to engage in additional tasks as a means of increasing access to situation-model cues. However, even when learners have enhanced access to such cues, they might not use them if they do not understand that this is the level of mental text representation that will be assessed during testing. Hence, the comprehension-test-expectancy intervention is another effective approach to improving relative metacomprehension accuracy because it guides learners toward selecting situation-model cues: Learners are instructed that upcoming tests will assess their deeper understanding and receive respective sample test questions (Griffin et al. 2019b; Thiede et al. 2011; Wiley et al. 2008, 2016a). For example, in a study by Griffin et al. (2019b, Experiment 1), learners in the memory-test-expectancy group were told that their memory of specific details would be tested, and they completed sample memory test questions about practice texts. In contrast, learners in the comprehension-test-expectancy group were told that their ability to draw inferences would be tested, and they completed sample inference test questions about practice texts. Learners in the no-expectancy group were only told that they would be taking tests and did not receive any sample test questions. Concerning the critical texts, all learners completed both memory and inference tests. The results revealed that the learners in the no-expectancy and memory-test-expectancy groups made judgments that were more predictive of their performance on the memory than the inference tests (i.e., greater relative metamemory accuracy). However, the learners in the comprehension-test-expectancy group made judgments that were more closely related to their performance on the inference than the memory tests (i.e., greater relative metacomprehension accuracy).

Prior research has indicated that learners typically view the concept of text comprehension in terms of memory of textual information rather than deep understanding and hence anticipate tests requiring recall (see, e.g., Wiley et al. 2005). Consequently, learners tend to rest upon memory-based instead of situation-model cues to evaluate their comprehension (Jaeger and Wiley 2014, Experiment 2; Thiede et al. 2010, Experiment 1). Thus, poor relative metacomprehension accuracy at least in part results from learners' incorrect assumptions about what comprehension means and what they therefore will be tested on. The comprehension-test-expectancy intervention is useful to override this reading-for-memory mindset so that learners select appropriate cues and make accurate judgments (see, e.g., also Griffin et al. 2019a; Wiley et al. 2016b). The test-expectancy effect even emerged when the intervention was implemented after reading (Griffin et al. 2019b, Experiment 3). This result confirms that test expectancies influence which cues learners select when making judgments rather than learners' access to particular cues. Moreover, when combined with self-explaining, the comprehension-test-expectancy intervention provided a unique, nonoverlapping benefit for relative metacomprehension accuracy (Griffin et al. 2019b, Experiment 4; see also Wiley et al. 2008, Experiment 2). This finding supports the theoretical distinction between interventions that promote cue generation versus cue selection.

The Present Study

The reviewed literature shows that interventions aimed at increasing the use of situation-model cues usually support relative metacomprehension accuracy. However, the extent of their effectiveness is unclear. Therefore, we conducted a meta-analysis examining the overall effectiveness of situation-model-approach interventions. In doing so, we focused on the most thoroughly investigated interventions described above, namely delayed-summary writing, delayed-keywords listing, delayed-diagram completion, self-explaining, concept mapping, rereading, and setting a comprehension-test expectancy. We expected to find a positive average effect across all interventions (Hypothesis 1) as well as for each intervention individually (Hypothesis 2).

The different interventions come with different costs and benefits. For example, in contrast to interventions that are executed directly during reading, such as self-explaining, delayed-generation tasks have to be performed several minutes after reading to yield positive effects. Moreover, whereas some interventions can easily be implemented in instruction, such as rereading, others, for instance, concept mapping, require training before learners can carry them out in a way that leads to positive effects. In addition, some interventions, such as setting a comprehension-test expectancy, do not facilitate external diagnostic processes, but some of them, like summary writing, produce outputs that enable instructors to evaluate learners' understanding. Yet, the effectiveness of the interventions relative to each other is unknown. Thus, in a further step, we statistically compared their effectiveness (Research Question 1).

Although all of the situation-model-approach interventions aim at enhancing learners' use of situation-model cues, they do so in different ways. Specifically, they can be classified into interventions promoting the generation of, attention to, or selection of situation-model cues (see, e.g., Griffin et al. 2013; Wiley et al. 2016b). This theoretical distinction has been empirically supported as the effects of self-explaining (cue-generation intervention) were nonoverlapping with the effects of rereading (cue-attention intervention) and setting a comprehension-test expectancy (cue-selection intervention; Griffin et al. 2008, Experiment

2; Griffin et al. 2019b, Experiment 4; Wiley et al. 2008, Experiment 2). Furthermore, Thiede et al. (2019) suggested to distinguish between encoding-based and retrieval-based cue-generation interventions. Some cue-generation interventions, namely self-explaining and concept mapping, encourage learners to *encode* a text in a manner that supports the construction of a situation model and thus the availability of situation-model cues, whereas other cue-generation interventions, namely delayed-generation tasks, promote learners to *retrieve* information about their situation model after reading and hence the availability of situation-model cues. As it is unclear whether the different intervention types (cue-generation encoding-based interventions, cue-generation retrieval-based interventions, cue-attention intervention, and cue-selection intervention) vary in their effectiveness for supporting relative metacomprehension accuracy, we explored to what extent their effects differ (Research Question 2).

Finally, although not the focus of our study, we were interested in the impact of the interventions not only on relative metacomprehension accuracy but also on comprehension (Research Question 3). To draw conclusions about the overall effectiveness of the interventions for supporting self-regulated learning from texts, both their indirect effects via metacomprehension and their direct effects on comprehension need to be considered.

Method

Literature Search and Study Selection

In a recent meta-analysis, we investigated learners' baseline level of relative metacomprehension accuracy (Prinz et al. 2020). In the context of this recent meta-analysis, we conducted a comprehensive literature search for studies on relative metacomprehension accuracy, yielding 800 unique records. After a first screening of these records for relevance to the study aim, 214 records remained. Then, the following inclusion criteria were applied:

- The focus was learning from texts (i.e., not learning of single sentences or vocabularies, more than one text).
- Metacomprehension judgments were provided before testing (i.e., predictions).
- Relative metacomprehension accuracy was operationalized with the Gamma or Pearson correlation.
- There was a sample or condition in which participants did not receive an intervention or any other kind of manipulation (e.g., prejudgment recall of the texts, sample questions about the critical texts, monetary incentives to provide accurate judgments).
- Sufficient statistical information for effect size computation was reported.

In addition, the following exclusion criteria were set:

- The data were also reported in another record (i.e., when a study was published in a journal as well as in a conference proceeding or dissertation, only the journal article was included).
- Participants were learners with learning disabilities or second-language learners.
- The authors of the meta-analysis did not master the language in which the study was reported (e.g., Japanese).
- The texts were presented in the context of a multimedia learning environment or illustrations or animations were presented along with the texts.

- Participants had the opportunity to regulate their learning before testing.
- Paradigm-atypical reading procedures, predictions, or tests were applied (e.g., participants were allowed to highlight text, participants made their predictions 1 day after reading the texts, the texts were available during testing).

After this second screening, 66 eligible records remained (see Prinz et al. 2020, for more details). For the present meta-analysis, we used this pool to further filter out relevant studies. To do so, additional inclusion criteria were applied:

- Learners were school or university students. Prior research has shown that, although school students' baseline level of relative metacomprehension accuracy is usually lower than that of university students (e.g., De Bruin et al. 2011; Redford et al. 2012, Experiment 2; Thiede et al. 2012, Study 1; see also Prinz et al. 2020), interventions that can improve relative metacomprehension accuracy among university students also have the potential to do so among school students (e.g., De Bruin et al. 2011; Redford et al. 2012, Experiment 2; see also Thiede et al. 2017). Therefore, we focused on learners at all education levels.
- The texts were expository texts.
- Comprehension of each text was assessed via questions that required inferences or application of the textual information (because the interventions are targeted at this kind of knowledge, that is, the situation model).
- A control condition in which participants did not receive an intervention was compared to a condition in which participants received a situation-model-approach intervention (i.e., an intervention that is implemented with the goal to focus learners on situation-model cues).
- For each particular intervention (i.e., with a similar implementation), more than one study had to meet the criteria for inclusion described so far. This led to the exclusion of the following interventions: generating drawings (Schleinschok et al. 2017; Thiede et al. 2019; but see Van de Pol et al. 2020, for an overview on the impact of drawing and mapping activities), generating explanations for others (Fukaya 2013), generating and answering questions (Bugg and McDaniel 2012), receiving advanced organizers (Dunlosky et al. 2002), reading texts with analogies (Wiley et al. 2018), and completing sentence-sorting tasks (Thomas and McDaniel 2007). The following interventions remained for inclusion: delayed-summary writing, delayed-keywords listing, delayed-diagram completion, self-explaining, concept mapping, rereading, and setting a comprehension-test expectancy.
- The typical implementation of the intervention was used. With regard to the *delayed-generation interventions*, the critical feature was that participants had to read all texts before being required to generate summaries, keywords, or diagrams. Although it was not a requirement for inclusion, it also became apparent that, in all studies on delayed-keywords listing identified for this meta-analysis, the participants received example keywords (a minor deviation occurred in the study by Engelen et al. 2018, Experiment 1, in which the participants not only received two example keywords but also had to list three more keywords themselves). Moreover, in all studies on delayed-diagram completion, the participants received a brief training on how to complete diagrams. With respect to delayed-summary writing, besides the study by Engelen et al. (2018, Experiment 1), in which the participants were asked to complete a two-sentence summary for a short practice text, in all other studies, the participants did not receive examples or training. The *self-explanation intervention* had to require participants to engage in self-explaining during reading. Additionally, in all studies

included, the participants were provided with a short example text along with potential self-explanation statements. Similarly, the requirement for the *concept-mapping intervention* was that participants had to construct concept maps during reading. In addition, in all studies, the participants received extensive instruction and training on constructing concept maps. The *rereading intervention* simply had to be implemented such that participants read a text twice before providing the prediction for that text. Finally, the critical feature with regard to the *comprehension-test-expectancy intervention* was that participants' knowledge about the upcoming comprehension tests had to be manipulated via both (a) an explicit description of the nature of the test questions and (b) sample test questions about practice texts.

As indicated previously, an inclusion criterion was that the control condition received no intervention. In the studies on delayed-diagram completion, after reading the texts, the participants in the control group completed a picture-matching task, which required them to find the differences between two pictures. As this filler task is very unlikely to affect learning processes, the studies were deemed appropriate for inclusion. Note, moreover, that in the studies on test expectancy, the participants in the control group neither received information about the nature of the upcoming test questions nor sample test questions.

- Sufficient statistical information for effect size computation was available. If the required information was not reported and not received through a previous author request, we asked the authors to provide the missing information. When this was unsuccessful, the study could not be included.

After applying these inclusion and exclusion criteria, a total of 28 effect sizes from 17 records remained. The studies were published between 2000 and 2020 and comprised a total of 2,236 participants. The flowchart in Fig. 1 provides a visual overview of the selection process. Table 1 presents the studies with corresponding effect size information for each intervention.

Coding of Studies

All studies were double coded by a trained research assistant and the first author of this meta-analysis. Discrepancies were resolved through discussion. In this way, an agreement rate of 100% was achieved.

Outcome Measures The main outcome measure was relative metacomprehension accuracy computed as an intraindividual correlation. To remain as consistent as possible in terms of the outcome measure used, we focused on Gamma correlations and sought to avoid Pearson correlations. When a study did not provide Gamma correlations, we first checked whether the information could be identified in the reviews by Thiede et al. (2009) or Wiley et al. (2016b) before making author requests.

As we were also interested in the impact of the interventions on comprehension, we additionally coded this outcome measure. Specifically, we focused on comprehension computed as the mean performance across the comprehension tests per participant. When a study did not provide the mean but the median test performance, we made author requests.

Moderator Variables For each study, we coded the specific *intervention* under investigation. Hence, the intervention was a moderator variable with seven categories: delayed-summary

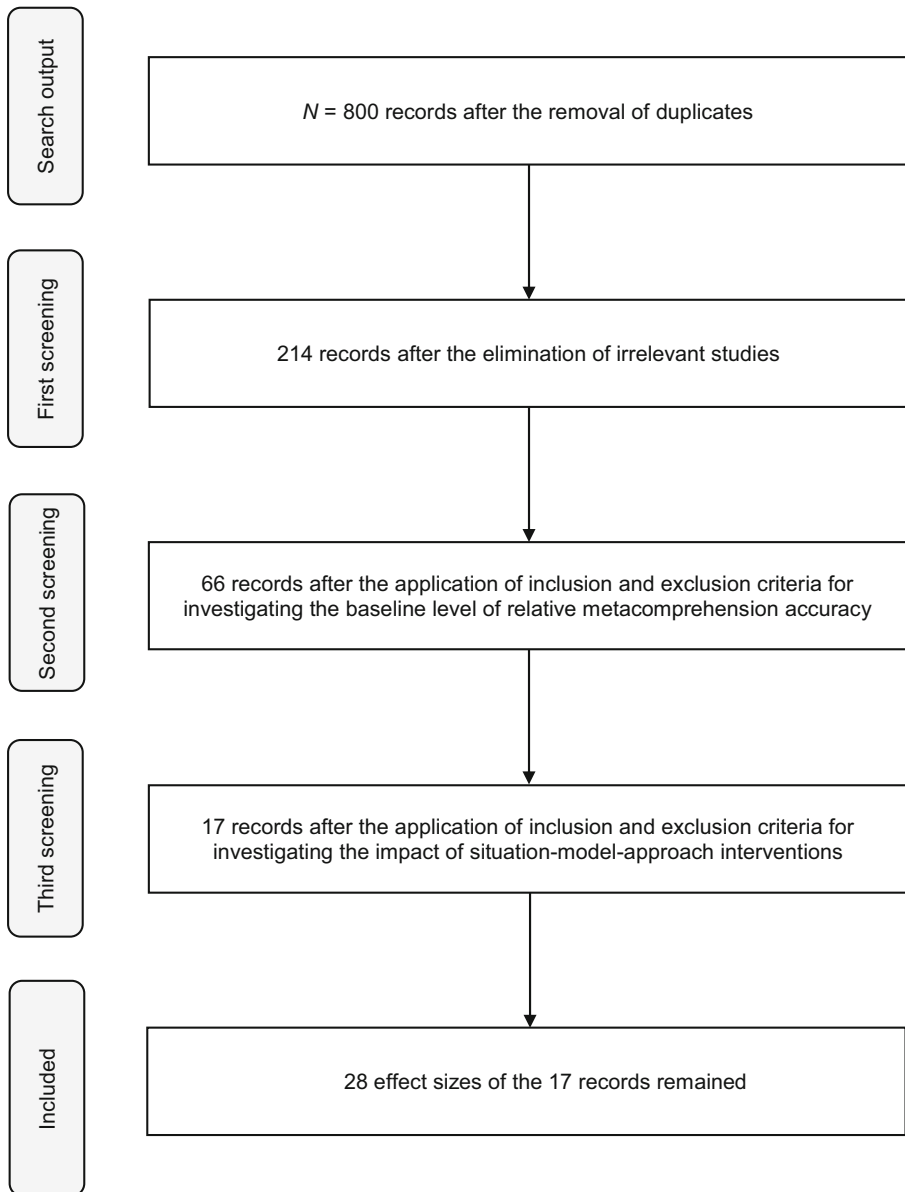


Fig. 1 Flowchart of the literature search

writing, delayed-keywords listing, delayed-diagram completion, self-explaining, concept mapping, rereading, and setting a comprehension-test expectancy.

Based on this moderator variable, we constructed a second moderator variable that reflected the *type of intervention* and had four categories: (a) cue-generation encoding-based intervention (i.e., self-explaining and concept mapping), (b) cue-generation retrieval-based intervention (i.e., delayed-summary writing, delayed-keywords listing, and delayed-diagram completion),

Table 1 Overview of the studies with effect size information regarding relative metacomprehension accuracy

Intervention	Study; subgroup if applicable	Intervention condition		Control condition		Effect size <i>g</i>	95% CI	
		<i>M</i>	<i>n</i>	<i>M</i>	<i>n</i> ^a		<i>LL</i>	<i>UL</i>
Delayed-summary writing	Anderson and Thiede (2008)	0.64	75	0.20	(75)	0.72	0.47	0.97
	Engelen et al. (2018, Exp. 1)	0.05	73	0.17	75	-0.18	-0.50	0.14
	Thiede and Anderson (2003, Exp. 1)	0.60	24	0.28	24	0.59	0.02	1.16
	Thiede and Anderson (2003, Exp. 2)	0.55	30	0.23	30	0.67	0.15	1.18
	Thiede et al. (2010, Exp. 1); at-risk college students ^b	0.48	32	0.08	(32)	0.57	0.20	0.94
	Thiede et al. (2010, Exp. 1); normal college students	0.63	74	0.21	(74)	0.60	0.35	0.84
Delayed-keywords listing	De Bruin et al. (2011, Exp. 1)	0.27	46	-0.01	48	0.45	0.05	0.86
	De Bruin et al. (2011, Exp. 2); fourth graders	0.11	46	0.18	39	-0.10	-0.52	0.32
	De Bruin et al. (2011, Exp. 2); sixth graders	0.42	33	0.00	33	0.72	0.22	1.21
	Engelen et al. (2018, Exp. 1)	0.09	81	0.17	75	-0.13	-0.44	0.19
	Thiede et al. (2003)	0.58	21	0.28	20	0.69	0.08	1.31
	Thiede et al. (2017)	0.41	175	0.04	174	0.47	0.26	0.68
	Waldeyer and Roelle (2020)	0.43	34	0.11	31	0.47	0.04	0.90
Delayed-diagram completion	Van de Pol et al. (2019, Exp. 1)	0.43	67	-0.00	64	0.69	0.34	1.04
	Van Loon et al. (2014)	0.56	40	0.07	34	0.75	0.28	1.21
Self-explaining	Griffin et al. (2019b, Exp. 4)	0.46	40	0.08	40	0.89	0.44	1.35
	Jaeger and Wiley (2014, Exp. 2)	-0.05	23	0.18	24	-0.32	-0.89	0.25
Concept mapping	Redford et al. (2012, Exp. 2) ^c	0.41	16	0.04	24	0.43	-0.20	1.05
	Thiede et al. (2010, Exp. 2)	0.66	19	0.30	(19)	0.51	0.05	0.97
Rereading	Chiang et al. (2010)	0.18	67	0.19	65	-0.01	-0.34	0.34
	Griffin et al. (2008, Exp. 1) ^d	0.46	36	0.22	36	0.48	0.02	0.94
	Griffin et al. (2008, Exp. 2) ^d	0.39	30	0.21	30	0.40	-0.10	0.91
	Margolin and Snyder (2018) ^e	0.29	64	0.37	(64)	-0.11	-0.40	0.18
	Rawson et al. (2000, Exp. 1)	0.57	36	0.24	39	0.59	0.13	1.05
	Rawson et al. (2000, Exp. 2)	0.55	26	0.19	30	0.81	0.27	1.35
Test expectancy	Griffin et al. (2019b, Exp. 1)	0.53	40	0.16	40	0.76	0.31	1.21
	Griffin et al. (2019b, Exp. 3)	0.46	24	0.08	24	0.79	0.22	1.37
	Griffin et al. (2019b, Exp. 4) ^f	0.51	40	0.08	40	0.95	0.49	1.41

^a When *n* for the control condition is reported in parentheses, this indicates that the study applied a within-subjects design and the sample size was already reported for the intervention condition

^b We extracted the effect size information for college students required to take a remedial reading course from this study as well because the fact that they accomplished to get into college warrants that they had sufficiently high capabilities to potentially benefit from the intervention

^c The data reported here differ somewhat from the original publication due to a data loss. However, the pattern of the results is the same

^d The Pearson correlation is reported and used

^e For calculating the variance of this effect, the correlation between relative metacomprehension accuracy in the two conditions had to be estimated as it was not reported. We set it at a medium correlation ($r = .30$), yielding a conservative estimate of the variance

^f The studies by Wiley et al. (2008) were not included because the data from this study were fed into the publication by Griffin et al. (2019b; J. Wiley, personal communication, November 11, 2019)

(c) cue-attention intervention (i.e., rereading), and (d) cue-selection intervention (i.e., setting a comprehension-test expectancy).

Effect Size Computation

We used Hedges's g (Hedges 1981) as the effect size measure. By incorporating a correction factor, Hedges's g provides a more conservative effect size estimate for small sample sizes compared to Cohen's d (see, e.g., Borenstein et al. 2009). Interpretative benchmarks according to Cohen (1988) were used: $g = 0.20$ indicates a small effect, $g = 0.50$ a medium effect, and $g = 0.80$ a large effect.

Procedure and Analysis

All statistical analyses were conducted with the Comprehensive Meta-Analysis Program (Borenstein et al. 2013). An alpha level of .05 was used for the analyses unless stated otherwise.

Average Effect To examine the impact of the situation-model-approach interventions, we computed the average effect and tested whether it differed significantly from zero. In doing so, each effect size was weighted by the inverse of its variance and a random-effects model was used. A random-effects model was preferred to a fixed-effect model because the selected studies are no exact replications but differ with regard to sample, design, materials, and procedure. Therefore, the true effect size likely varies across studies rather than being fixed. Specifically, the application of a random-effects model implies that the observed effect sizes are assumed to differ from the average effect due to sampling error (within-study variance) and a true-variation component (between-studies variance; see, e.g., Hunter and Schmidt 2015; Lipsey and Wilson 2001). The method of moments was used to estimate the variance between studies (see, e.g., Borenstein et al. 2009).

From Experiment 1 by Thiede et al. (2010), we extracted effect size information for two subgroups (i.e., college students required and college students not required to take a remedial reading course; see also footnote b for Table 1). Similarly, Experiment 2 by De Bruin et al. (2011) provided effect size information for two subgroups (i.e., fourth graders and sixth graders). As the subgroups contributed independent information, we used subgroups within studies instead of studies as the unit of analysis (see, e.g., Borenstein et al. 2009).

Moreover, Experiment 4 by Griffin et al. (2019b) reported data on two interventions, self-explaining and setting a comprehension-test expectancy, with the same control group serving as the point of comparison. Hence, to avoid problems with multiple comparisons (i.e., the lack of independence between the effect for self-explaining versus control and the effect for test expectancy versus control), we created a composite score, namely the mean of the effect for self-explaining versus control and the effect for test expectancy versus control. The variance of this composite score was computed based on the variance of each effect size and the correlation between the two effects (which was set to $r = .50$ given the equal sample size in conditions; see, e.g., Borenstein et al. 2009). The same issue applied to Experiment 1 by Engelen et al. (2018), which compared delayed-summary writing and delayed-keywords listing to the same control group. Again, we calculated a composite score that was the mean of the effect for delayed-summary writing versus control and the effect for delayed-keywords listing versus control. The composite score's variance was computed

based on the variance of each effect size and the correlation between the two effects (which was set to $r = .50$ given the approximately equal sample size in conditions; see, e.g., Borenstein et al. 2009).

Publication Bias We used three indicators to check for publication bias. We again applied a random-effects model. First, we generated a funnel plot. The effect size was mapped on the x -axis and the standard error on the y -axis. In the absence of a publication bias, the effect sizes would be distributed symmetrically around the average effect. In the presence of a publication bias, there would be some missing values near the bottom of the plot where studies with small samples are allocated (they generally have a larger standard error). Specifically, there would be a gap on the left side where the studies with small, nonsignificant effects would have been plotted if they had been published (see, e.g., Borenstein et al. 2009).

Second, we conducted Rosenthal's (1979) fail-safe N test. This test indicates how many unpublished zero-effect studies one would need to retrieve and incorporate into the meta-analysis before the p value for the average effect became nonsignificant.

Third, we conducted a trim-and-fill analysis (Duval and Tweedie 2000a, b). This analysis first trims asymmetric effect sizes from the right side of the funnel plot to calculate an unbiased estimate of the average effect. Then, to correct the variance of the effects, the plot is filled by reinserting the trimmed effect sizes on the right side and by adding their imputed counterparts on the left side.

Moderator Analyses We conducted moderator analyses to compare the effectiveness of the interventions. First, we tested for significant heterogeneity in effect sizes by calculating the Q statistic. Q reflects the total amount of study-to-study variation observed and can be compared to the expected variation (under the assumption that all variation is due to within-study error) to test the null hypothesis that all studies share a common true effect size (see, e.g., Borenstein et al. 2009; Cochran 1954). Additionally, we calculated the I^2 statistic. I^2 reflects the degree of heterogeneity, that is, what proportion of the total observed variation can be attributed to true variation in effect sizes (see, e.g., Higgins et al. 2003; Huedo-Medina et al. 2006). I^2 ranges from 0 to 100%. Higgins et al. (2003) suggest that values around 25%, 50%, and 75% may be interpreted as low, moderate, and high, respectively.

The categorical moderator variables were analyzed with special analysis of variance for meta-analyses. In doing so, mixed-effects models were applied. In mixed-effects models, the average effect within each moderator category (i.e., each particular intervention or intervention type) is calculated using a random-effects model and the difference across categories is assessed using a fixed-effect model (i.e., the categories are assumed to be exhaustive rather than sampled at random; see, e.g., Borenstein et al. 2009). In each moderator analysis, we pooled the within-category estimates of the variance of true effect sizes and applied this common estimate to all studies to obtain a more accurate estimate of the between-studies variance. As an effect size measure, we computed R^2 , which reflects the proportion of true variation among studies explained by a moderator. We applied the interpretative thresholds provided by Cohen (1988): $R^2 = .01$ denotes a small effect, $R^2 = .09$ a medium effect, and $R^2 = .25$ a large effect. Furthermore, the Q statistic was used to determine whether any two moderator categories significantly differed from each other (see, e.g., Borenstein et al. 2009). As already stated, based on Experiment 4 by Griffin et al. (2019b), effect sizes for self-explaining and setting a comprehension-test expectancy were calculated using the same control group. Likewise, based on Experiment 1 by Engelen et al. (2018), effect sizes for delayed-

summary writing and delayed-keywords listing were computed using the same control group. In the moderator analyses, the two effects from one study were treated as independent effects so that they could be included in the study pool for both interventions.

Results

The Average Effect of Situation-Model-Approach Interventions on Relative Metacomprehension Accuracy

We calculated the weighted mean correlation in the control and intervention conditions, respectively. There was a small average correlation in the control condition, $r = .14$, 95% CI [.09, .20], $k = 26$, and a medium-to-large average correlation in the intervention condition, $r = .41$, 95% CI [.33, .48], $k = 26$. Accordingly, the average effect across all situation-model-approach interventions on relative metacomprehension accuracy was significantly positive and of medium size, $g = 0.46$, 95% CI [0.33, 0.60], $p < .001$, $k = 26$ (as described in the [Procedure and Analysis](#) section, there were 26 effect sizes for this analysis because the two effects from Griffin et al. 2019b, Experiment 4, and Engelen et al. 2018, Experiment 1, were combined into one composite score, respectively; see Table 1 for the effect size information concerning the individual studies). Hence, in accordance with Hypothesis 1, situation-model-approach interventions are generally effective in improving relative metacomprehension accuracy.³

The set of effect sizes was heterogeneous, $Q(25) = 70.46$, $p < .001$, $I^2 = 65\%$. The I^2 statistic indicated that 65% of the observed variation could be attributed to true variation among studies, which reflects moderate-to-high heterogeneity (Higgins et al. 2003). Thus, we conducted a moderator analysis to compare the effectiveness of the particular interventions.

Publication Bias

First, the funnel plot in Fig. 2 shows that the spread of effect sizes about the average effect is not perfectly symmetrical. Specifically, toward the bottom of the plot, there is a gap on the left side, suggesting that some studies with small, nonsignificant effects are missing. This indicates that a publication bias might be present. Second, Rosenthal's (1979) fail-safe N test revealed that 843 studies with an effect size of zero would be needed to render the average effect nonsignificant. According to a guideline by Rosenthal (1991), as the fail-safe N exceeds $5 * k + 10$ ($5 * 26 + 10 = 140 < 843$), the average effect seems to be robust to a potential publication bias. Third, the trim-and-fill analysis imputed four studies. However, the estimate of the unbiased average effect was $g = 0.41$, 95% CI [0.28, 0.54], which would represent only a small change in the average effect size. In conclusion, the impact of publication bias in the present study seems to be negligible.

³ As we included studies with both school and university students, we checked whether the average effect differed between the two groups of learners. A moderator analysis revealed that the average effects for school students, $g = 0.39$, 95% CI [0.15, 0.63], $k = 8$, and university students, $g = 0.50$, 95% CI [0.33, 0.66], $k = 18$, did not significantly differ from each other, $Q(1) = 0.52$, $p = .473$, $R^2 < .01$.

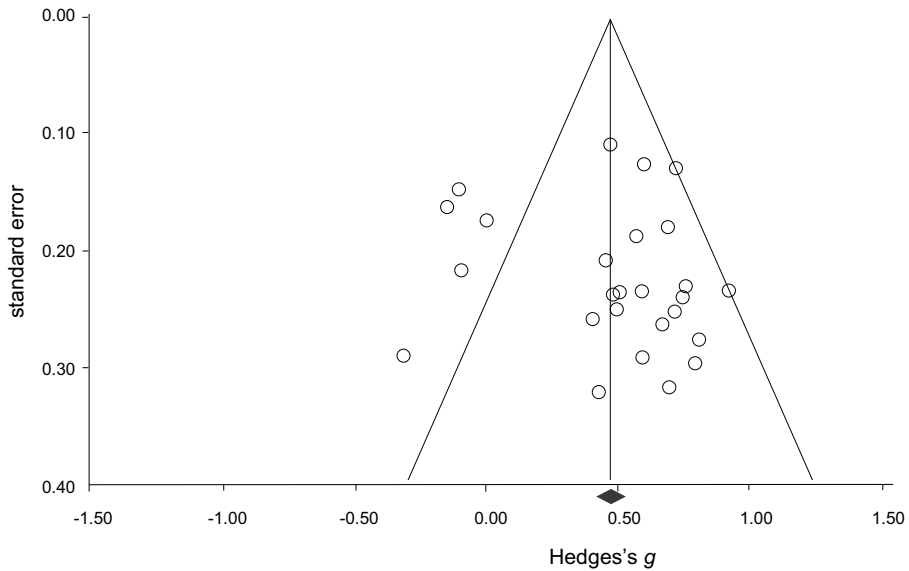


Fig. 2 Funnel plot

Comparisons of the Interventions with Regard to Relative Metacomprehension Accuracy

Interventions The moderator analysis with intervention revealed a significant positive average effect of small-to-medium size for rereading and delayed-keywords listing, of medium size for delayed-summary writing, of medium-to-large size for delayed-diagram completion, and of large size for setting a comprehension-test expectancy. In addition, there was a marginally significant positive average effect of medium size for concept mapping ($p = .097$). The average effect for self-explaining was not significant (see Table 2). Hence, apart from self-explaining, the results are in line with Hypothesis 2 in that each particular intervention supports relative metacomprehension accuracy.

The moderator analysis was not significant, $Q(6) = 5.71, p = .456, R^2 = .03$. Pairwise comparisons with Bonferroni correction of the alpha level (corrected alpha = .002) showed that none of the differences between the interventions reached the set level of significance (see

Table 2 Effect size and heterogeneity statistics for the interventions with regard to relative metacomprehension accuracy

Intervention	Number of effect sizes <i>k</i>	Effect size <i>g</i>	95% CI		<i>Q</i>	<i>p</i>	<i>I</i> ²
			<i>LL</i>	<i>UL</i>			
Delayed-summary writing	6	0.49	0.22	0.77	21.46	.001	77
Delayed-keywords listing	7	0.34	0.08	0.61	18.23	.001	67
Delayed-diagram completion	2	0.72	0.22	1.21	0.04	.851	< 1
Self-explaining	2	0.34	-0.20	0.89	10.68	.001	91
Concept mapping	2	0.47	-0.08	1.03	0.04	.841	< 1
Rereading	6	0.32	0.02	0.61	15.45	.009	68
Test expectancy	3	0.84	0.40	1.27	0.39	.825	< 1

Table 3 Pairwise comparisons of the average effects of the interventions with regard to relative metacomprehension accuracy

Intervention	1.		2.		3.		4.		5.		6.		7.	
	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>
1. Delayed-summary writing	–													
2. Delayed-keywords listing	0.51	.474	–											
3. Delayed-diagram completion	0.65	.421	2.11	.147	–									
4. Self-explaining	0.19	.666	0.01	.963	0.55	.458	–							
5. Concept mapping	0.01	.965	0.21	.649	0.96	.328	0.06	.800	–					
6. Rereading	0.60	.437	0.02	.891	2.17	.141	0.01	.986	0.27	.605	–			
7. Test expectancy	1.94	.164	4.70	.030	0.41	.523	1.31	.252	2.32	.128	4.79	.029	–	

Table 3). However, the average effect for setting a comprehension-test expectancy tended to be larger than the average effects for delayed-keywords listing and rereading ($ps < .05$). Therefore, with respect to Research Question 1, the results indicate that the interventions do not substantially differ with regard to their impact on relative metacomprehension accuracy. Yet, there is a tendency for the test-expectancy intervention to be particularly effective, especially compared with delayed-keywords listing and rereading.

Concerning delayed-summary writing, delayed-keywords listing, self-explaining, and rereading, the test of heterogeneity was significant (see Table 2). For delayed-keywords listing, there was some variation with regard to learners’ education level. However, an explorative moderator analysis revealed that, although the average effect was only marginally significant for school students, $g = 0.27$, 95% CI $[-0.03, 0.57]$, $k = 5$, it did not significantly differ from the average effect for university students, $g = 0.58$, 95% CI $[0.03, 1.14]$, $k = 2$; $Q(1) = 0.94$, $p = .333$, $R^2 < .01$. We refrained from conducting further moderator analyses within each intervention because there were only few studies for each, causing too little variability in the data (i.e., few cases per moderator category; e.g., there was only one study with school students for delayed-summary writing and none for self-explaining and rereading).

Intervention Types The moderator analysis with intervention type revealed a significant positive average effect of small-to-medium size for the cue-generation encoding-based interventions and the cue-attention intervention, of medium size for the cue-generation retrieval-based interventions, and of large size for the cue-selection intervention (see Table 4).

The moderator analysis was also not significant, $Q(3) = 4.07$, $p = .254$, $R^2 = .10$. Pairwise comparisons with Bonferroni correction of the alpha level (corrected alpha = .008) showed that none of the differences between the intervention types reached the set level of significance (see Table 5). However, in accordance with the analysis on the seven specific interventions, the average effect for the cue-selection intervention tended to be larger than the average effect for the cue-attention intervention ($p = .029$). Hence, concerning Research Question 2, it can be concluded that the different intervention types do not substantially differ in their effectiveness for supporting relative metacomprehension accuracy. Yet, the cue-selection intervention of setting the right test expectancy tends to be more effective than the cue-attention intervention, that is, rereading.

Apart from the cue-selection intervention, the test of heterogeneity for the other intervention types was significant (see Table 4). For the cue-generation retrieval-based interventions, there was

Table 4 Effect size and heterogeneity statistics for the intervention types with regard to relative metacomprehension accuracy

Intervention	Number of effect sizes <i>k</i>	Effect size <i>g</i>	95% CI		<i>Q</i>	<i>p</i>	<i>I</i> ²
			<i>LL</i>	<i>UL</i>			
Cue-generation encoding-based interventions	4	0.41	0.03	0.79	10.77	.013	72
Cue-generation retrieval-based interventions	15	0.45	0.28	0.63	46.54	< .001	70
Cue-attention intervention	6	0.31	0.03	0.60	15.45	.009	68
Cue-selection intervention	3	0.84	0.41	1.26	0.39	.825	< 1

some variation regarding learners' education level. An explorative moderator analysis revealed that, although both average effects were significant, the one for school students, $g = 0.31$, 95% CI [0.10, 0.52], $k = 8$, was marginally significantly smaller than the one for university students, $g = 0.62$, 95% CI [0.39, 0.86], $k = 7$; $Q(1) = 3.77$, $p = .052$, $R^2 = .27$. We did not run further moderator analyses within each intervention type because there were only few studies for each, restricting the variability in the data (i.e., few cases per moderator category).

The Average Effect of Situation-Model-Approach Interventions on Comprehension

The average effect across the situation-model-approach interventions on comprehension was significantly positive but small, $g = 0.13$, 95% CI [0.002, 0.26], $p = .046$, $k = 26$ (see Table 6 for the effect size information concerning the individual studies). Thus, with respect to Research Question 3, the results suggest that situation-model-approach interventions generally improve not only metacomprehension but also comprehension although to a lesser extent.⁴

The set of effect sizes was heterogeneous with moderate-to-high heterogeneity, $Q(25) = 72.74$, $p < .001$, $I^2 = 66\%$ (Higgins et al. 2003). Hence, we next compared the effectiveness of the particular interventions in a moderator analysis.

Comparisons of the Interventions with Regard to Comprehension

Interventions The moderator analysis with intervention showed that there was a significant medium-to-large positive average effect for concept mapping and a marginally significant small-to-medium positive average effect for the test-expectancy intervention ($p = .054$). None of the other interventions yielded a significant average effect (see Table 7).

The moderator analysis was significant, $Q(6) = 15.57$, $p = .016$, $R^2 = .37$. Pairwise comparisons with Bonferroni correction of the alpha level (corrected alpha = .002) showed that the average effect for delayed-summary writing was significantly smaller than the average effects for concept mapping and setting a comprehension-test expectancy and by trend also smaller than the average effect for self-explaining ($p = .015$). Moreover, the average effect for delayed-keywords listing tended to be smaller than the average effects for concept mapping and setting a comprehension-test expectancy ($p < .05$; see Table 8). Hence, with respect to Research

⁴ Again, we checked whether the average effect differed between school and university students. A moderator analysis revealed that, whereas the average effect was significant for university students, $g = 0.16$, 95% CI [0.01, 0.33], $k = 18$, it was not significant for school students, $g = 0.07$, 95% CI [-0.16, 0.29], $k = 8$. However, this difference was not statistically significant, $Q(1) = 0.49$, $p = .483$, $R^2 < .01$.

Table 5 Pairwise comparisons of the average effects of the intervention types with regard to relative metacomprehension accuracy

Intervention	1.		2.		3.		4.	
	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>
1. Cue-generation encoding-based interventions	–							
2. Cue-generation retrieval-based interventions	0.04	.836	–					
3. Cue-attention intervention	0.08	.781	0.64	.423	–			
4. Cue-selection intervention	2.13	.145	2.98	.084	4.79	.029	–	

Question 3, the results further indicate that there is a tendency for concept mapping and the test-expectancy intervention to be particularly effective, whereas delayed-summary writing and delayed-keywords listing are rather ineffective for supporting comprehension.

Only for rereading was the test of heterogeneity significant (see Table 7). We refrained from conducting further moderator analyses within this intervention, however, because of too little variability in the data.

Intervention Types The moderator analysis with intervention type revealed that there was a significant medium positive average effect for the cue-generation encoding-based interventions and a marginally significant small-to-medium positive average effect for the cue-selection intervention ($p = .054$). The average effects for the other two intervention types were not significant (see Table 9).

The moderator analysis was significant, $Q(3) = 12.03$, $p = .007$, $R^2 = .37$. In line with the analysis on the seven specific interventions, pairwise comparisons with Bonferroni correction of the alpha level (corrected alpha = .008) showed that the average effect for the cue-generation retrieval-based interventions was significantly smaller than the average effect for the cue-generation encoding-based interventions and by trend also smaller than the average effect for the cue-selection intervention ($p = .017$; see Table 10). Thus, concerning Research Question 3, these results reinforce that the cue-generation retrieval-based interventions do not support comprehension, whereas the cue-generation encoding-based interventions and the cue-selection intervention of setting the right test expectancy are beneficial in this regard.

The test of heterogeneity was significant for the cue-generation retrieval-based interventions and the cue-attention intervention (see Table 9). Concerning the cue-generation retrieval-based interventions, we ran an explorative moderator analysis with learners' education level. The average effects for school students, $g = -0.03$, 95% CI $[-0.19, 0.14]$, $k = 8$, and university students, $g = -0.05$, 95% CI $[-0.24, 0.14]$, $k = 7$, were not significant and did not significantly differ from each other, $Q(1) = 0.03$, $p = .874$, $R^2 < .01$. We did not run further moderator analyses within each intervention type because there were only few studies for each, restricting the variability in the data (i.e., few cases per moderator category).

Discussion

The present meta-analysis investigated to what extent relative metacomprehension accuracy can be enhanced through interventions that promote learners' use of situation-model cues for

Table 6 Overview of the studies with effect size information regarding comprehension

Intervention	Study; subgroup if applicable	Intervention condition		Control condition		Effect size <i>g</i>	95% CI	
		<i>M</i>	<i>n</i>	<i>M</i>	<i>n</i> ^a		<i>LL</i>	<i>UL</i>
Delayed-summary writing	Anderson and Thiede (2008)	0.59	75	0.62	(75)	-0.12	-0.35	0.10
	Engelen et al. (2018, Exp. 1)	2.82	86	3.16	84	-0.43	-0.73	-0.13
	Thiede and Anderson (2003, Exp. 1)	0.74	25	0.71	25	0.15	-0.40	0.69
	Thiede and Anderson (2003, Exp. 2)	0.49	30	0.53	30	-0.28	-0.79	0.22
	Thiede et al. (2010, Exp. 1); at-risk college students ^b	4.21	32	4.02	(32)	0.14	-0.20	0.48
	Thiede et al. (2010, Exp. 1); normal college students	5.84	74	6.31	(74)	-0.32	-0.55	-0.09
Delayed-keywords listing	De Bruin et al. (2011, Exp. 1)	2.54	46	2.58	48	-0.04	-0.44	0.36
	De Bruin et al. (2011, Exp. 2); fourth graders ^c	2.09	46	1.92	39	0.17	-0.26	0.59
	De Bruin et al. (2011, Exp. 2); sixth graders ^c	2.65	33	2.36	33	0.29	-0.19	0.77
	Engelen et al. (2018, Exp. 1)	3.02	85	3.16	84	-0.17	-0.47	0.13
	Thiede et al. (2003)	0.44	22	0.47	22	-0.14	-0.72	0.44
	Thiede et al. (2017)	2.42	240	2.45	240	-0.02	-0.20	0.15
	Waldeyer and Roelle (2020)	11.31	36	9.76	36	0.48	0.02	0.95
Delayed-diagram completion	Van de Pol et al. (2019, Exp. 1)	1.17	85	1.20	84	-0.03	-0.33	0.28
	Van Loon et al. (2014)	49.22	44	44.19	39	0.30	-0.13	0.73
Self-explaining	Griffin et al. (2019b, Exp. 4)	0.51	40	0.46	40	0.55	0.11	0.99
	Jaeger and Wiley (2014, Exp. 2)	2.25	23	2.11	25	0.13	-0.43	0.69
Concept mapping	Redford et al. (2012, Exp. 2)	2.29	32	1.86	32	0.38	-0.11	0.87
	Thiede et al. (2010, Exp. 2)	3.75	21	2.69	(21)	0.99	0.48	1.49
Rereading	Chiang et al. (2010)	2.01	75	2.06	69	-0.10	-0.42	0.23
	Griffin et al. (2008, Exp. 1)	0.69	36	0.70	36	-0.05	-0.50	0.41
	Griffin et al. (2008, Exp. 2)	0.68	31	0.70	31	-0.11	-0.60	0.38
	Margolin and Snyder (2018)	0.66	64	0.60	(64)	< 0.01	-0.24	0.24
	Rawson et al. (2000, Exp. 1)	0.59	40	0.52	40	0.42	-0.02	0.86
	Rawson et al. (2000, Exp. 2)	0.86	30	0.70	30	1.34	0.78	1.89
Test expectancy	Griffin et al. (2019b, Exp. 1)	0.51	40	0.47	40	0.36	-0.08	0.80
	Griffin et al. (2019b, Exp. 3)	0.47	24	0.45	24	0.22	-0.34	0.78
	Griffin et al. (2019b, Exp. 4) ^d	0.51	40	0.46	40	0.46	0.02	0.90

^a When *n* for the control condition is reported in parentheses, this indicates that the study applied a within-subjects design and the sample size was already reported for the intervention condition

^b We extracted the effect size information for college students required to take a remedial reading course from this study as well because the fact that they accomplished to get into college warrants that they had sufficiently high capabilities to potentially benefit from the intervention

^c The mean across participants' median (not average) comprehension performance on the tests in each group is reported. As the data are not available anymore, the participants' average comprehension-test performance could not be computed

^d The studies by Wiley et al. (2008) were not included because the data from this study were fed into the publication by Griffin et al. (2019b; J. Wiley, personal communication, November 11, 2019)

judging their comprehension. In particular, these interventions were delayed-summary writing, delayed-keywords listing, delayed-diagram completion, self-explaining, concept mapping, rereading, and setting a comprehension-test expectancy.

Table 7 Effect size and heterogeneity statistics for the interventions with regard to comprehension

Intervention	Number of effect sizes <i>k</i>	Effect size <i>g</i>	95% CI		<i>Q</i>	<i>p</i>	<i>I</i> ²
			<i>LL</i>	<i>UL</i>			
Delayed-summary writing	6	−0.17	−0.39	0.05	9.26	.099	46
Delayed-keywords listing	7	0.06	−0.16	0.27	7.72	.260	22
Delayed-diagram completion	2	0.11	−0.28	0.50	1.44	.230	31
Self-explaining	2	0.37	−0.09	0.82	1.31	.253	24
Concept mapping	2	0.67	0.22	1.13	2.85	.091	65
Rereading	6	0.18	−0.06	0.42	24.33	<.001	80
Test expectancy	3	0.36	−0.01	0.72	0.44	.801	<1

Effectiveness of Situation-Model-Approach Interventions

By showing that situation-model-approach interventions generally improve relative metacomprehension accuracy, the results of this meta-analysis reinforce previous narrative reviews (e.g., Griffin et al. 2013, 2019a; Wiley et al. 2016b). In addition, the results unveil that the improvement is quite considerable (i.e., medium effect). Therefore, the interventions should be used in instruction to support the accuracy of learners' judgments about their comprehension and thus their self-regulated learning from texts.

With the exception of self-explaining, each of the investigated situation-model-approach interventions reliably enhances relative metacomprehension accuracy. Yet, the test-expectancy intervention seems to be particularly effective, especially in comparison to delayed-keywords listing and rereading. Accordingly, a further analysis on the intervention types showed that the cue-selection intervention of setting the appropriate test expectancy tends to be more beneficial than the cue-attention intervention of rereading. Hence, this meta-analysis extends prior research not only by revealing the overall degree of effectiveness of the interventions but also by indicating which intervention is comparatively more effective (i.e., test expectancy), less effective (i.e., rereading), or unreliable (i.e., self-explaining).

The average effect for self-explaining was heterogeneous and not significant. Of the two studies on self-explaining in the present meta-analysis, the effect was significantly positive in the study by Griffin et al. (2019b, Experiment 4) but not significant and even negative in the study by Jaeger and Wiley (2014, Experiment 2; see Table 1). As conducting further moderator analyses within this intervention was not feasible with two studies, we compared the studies on variables that could potentially have caused the discrepancy in effects (see Prinz et al. 2020). However, the studies were very similar with regard to sample, design, materials, and procedure so that differences in these aspects are unlikely to explain the variation among effects. In the study by Jaeger and Wiley (2014, Experiment 2), analyses of self-reported cue use showed that self-explaining did not increase learners' use of situation-model cues but actually tended to reduce the likelihood of using such cues compared to when learners did not engage in self-explaining. A possible reason for this outcome is that being required to read and simultaneously self-explain cognitively overwhelmed the learners, preventing them from generating or at least monitoring situation-model cues. Actually, a similar outcome was revealed for problem-solving tasks in previous research: In contrast to other generation activities, such as fully or partially generating the solution procedure by oneself, self-explaining the solution steps did not improve monitoring accuracy presumably because it was too cognitively demanding and thus led to low-quality explanations that provided invalid

Table 8 Pairwise comparisons of the average effects of the interventions with regard to comprehension

Intervention	1.		2.		3.		4.		5.		6.		7.	
	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>
1. Delayed-summary writing	–													
2. Delayed-keywords listing	3.33	.068	–											
3. Delayed-diagram completion	2.14	.144	0.11	.737	–									
4. Self-explaining	5.92	.015	2.76	.097	1.19	.275	–							
5. Concept mapping	12.59	<.001	8.03	.005	3.09	.079	0.73	.394	–					
6. Rereading	3.44	.064	0.48	.487	0.07	.787	0.15	.698	1.54	.214	–			
7. Test expectancy	9.85	.0017	5.04	.025	2.31	.129	0.01	.918	1.68	.195	0.25	.615	–	

Table 9 Effect size and heterogeneity statistics for the intervention types with regard to comprehension

Intervention	Number of effect sizes <i>k</i>	Effect size <i>g</i>	95% CI		<i>Q</i>	<i>p</i>	<i>I</i> ²
			<i>LL</i>	<i>UL</i>			
Cue-generation encoding-based interventions	4	0.52	0.20	0.84	5.42	.144	45
Cue-generation retrieval-based interventions	15	−0.03	−0.17	0.11	25.38	.031	45
Cue-attention intervention	6	0.18	−0.06	0.42	24.33	<.001	80
Cue-selection intervention	3	0.36	−0.01	0.72	0.44	.801	<1

cues (e.g., Baars et al. 2018; see also Van Gog et al. 2020). Hence, the benefit of self-explaining for relative metacomprehension accuracy might depend on learners’ available cognitive resources. In fact, in studies that applied a combined self-explanation and rereading intervention, the effect on relative metacomprehension accuracy was consistently positive (Griffin et al. 2008, Experiment 2; Wiley et al. 2008, Experiment 2). Self-explaining during rereading might ensure that learners can allocate ample resources to generating and monitoring situation-model cues. Therefore, an important direction for future research is to explore what condition, such as higher working-memory capacity, or what support, such as rereading instructions, is needed for the self-explanation intervention to reliably improve relative metacomprehension accuracy.

There was also significant heterogeneity among the effects for delayed-summary writing, delayed-keywords listing, and rereading. As there were rather few studies on each particular intervention with not much variability in methodological implementation, we again could not conduct moderator analyses to explain the variability in effects. Hence, more studies on these interventions including the examination of potential moderating influences are needed. For example, it seems likely that, as in the case of self-explaining, the effect of rereading depends on individual differences in cognitive capacity and reading skill. In fact, Griffin et al. (2008) found that rereading particularly benefited learners with low working-memory capacity or low reading skill (see also Ikeda and Kitagami 2013). Furthermore, the effects of delayed-summary writing and delayed-keywords listing might depend on the quality of the resulting products and learners’ cognitive capacity to produce them. For instance, if keywords do not tap central ideas from a text but rather unimportant facts, their benefit for yielding situation-model cues might be limited. This assumption is indirectly supported by the finding that, for the cue-generation retrieval-based interventions taken together (for these interventions jointly, there was sufficient variation regarding learners’ education level to conduct a moderator analysis), the average effect tended to be smaller for school students than for university students. Thus, when applying delayed-generation tasks to promote relative metacomprehension accuracy in

Table 10 Pairwise comparisons of the average effects of the intervention types with regard to comprehension

Intervention	1.		2.		3.		4.	
	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>	<i>Q</i>	<i>p</i>
1. Cue-generation encoding-based interventions	–							
2. Cue-generation retrieval-based interventions	12.08	.001	–					
3. Cue-attention intervention	1.23	.267	2.13	.144	–			
4. Cue-selection intervention	0.64	.423	5.68	.017	0.25	.615	–	

instruction, learners with lower cognitive capacity, especially younger learners, need to be supported in generating sophisticated products, for example, by receiving extensive training first.

Beyond the effects on relative metacomprehension accuracy, the effects on comprehension as well as practical considerations play important roles in determining the interventions' utility for supporting self-regulated learning from texts. Concerning comprehension, our results showed that, although the effect is comparatively small, situation-model-approach interventions are generally beneficial. Moreover, there is a tendency for concept mapping and setting a comprehension-test expectancy to be particularly effective, especially compared with delayed-generation tasks. Concerning practical utility, the test-expectancy intervention has the advantages that it does not require training, a delay after reading, or additional instruction time beyond reading, and it has the potential to improve learners' metacognitive knowledge. In fact, the last two aspects are unique benefits of the test-expectancy intervention. Whereas all other examined interventions require learners to engage in an additional task that has the by-product of improving relative metacomprehension accuracy, the test-expectancy intervention helps learners actively apply the appropriate learning goal to adjust their metacognitive processes as they monitor and judge their comprehension. An advantage of this approach is that learners can internalize this goal and adopt it in future learning episodes (cf. Griffin et al. 2019b). In contrast, rereading, for example, is easy to implement because it does not require training or a delay, but additional instruction time has to be spent on reading and no general metacognitive knowledge can be acquired. Given the overall picture, setting a comprehension-test expectancy appears to be a particularly beneficial intervention for enhancing self-regulated learning from texts.

Consequently, it seems important to shed more light on the way in which the test-expectancy intervention exerts its effects. Also in general, the mechanisms underlying the effectiveness of situation-model-approach interventions should be investigated more closely. For instance, at a more fine-grained level, the interventions might improve cue use by reducing cognitive load. This proposition is in line with the effort monitoring and regulation framework (De Bruin et al. 2020), which suggests that, for effective self-regulated learning, cognitive load needs to be optimized during and after the primary task. Cue-generation and cue-attention interventions might optimize cognitive load during learning, whereas cue-selection interventions might optimize cognitive load at the time of judgment. More precisely, cue-generation and cue-attention interventions might help learners devote cognitive resources to monitoring. For example, during rereading, the reduced need for comprehension-related processing (e.g., Millis et al. 1998) might free cognitive capacity for monitoring. Generating summaries, keywords, diagrams, or concept maps might promote monitoring by enabling cognitive offloading as written text and graphics provide an external representation of a learner's understanding (cf. Nückles et al. 2020). Concept mapping, for instance, yields a visual representation of a learner's situation model that can be reinspected during learning, which might ease working-memory capacity for monitoring. Moreover, concept mapping, self-explaining, and delayed-generation tasks simply provide additional opportunities to engage with the learning material on a situation-model level and hence to monitor the situation model. In contrast, the cue-selection intervention of setting the appropriate test expectancy might alleviate cognitive load at the time of judgment by helping learners focus on relevant cues. Learners may often be overwhelmed by the many available cues and hence rely on heuristic and memory-based cues because these cues are more salient and less effortful to use (cf. Griffin et al. 2009). When learners are informed about what comprehension entails and what

cues are therefore useful for evaluating their comprehension, however, they might be better able to disregard heuristic and memory-based cues and to concentrate on situation-model cues. In conclusion, our results indicate that optimizing cognitive load both during and after the primary task of comprehending a text promotes learners to more accurately judge their understanding. However, reducing the load after the primary task, that is, at the time of judgment, seems particularly effective.

The effort monitoring and regulation framework (De Bruin et al. 2020) also proposes that self-regulated learning can be impaired when learners misinterpret their invested mental effort. That is, although high mental effort is not necessarily indicative of poor learning, learners tend to judge their learning as low when experiencing high mental effort (e.g., Baars et al. 2014, 2018, Experiment 2; Dunlosky et al. 2006; Koriat et al. 2009; Schleinschok et al. 2017; see also Baars et al. 2020; Carpenter et al. 2020). Hence, because learners' invested mental effort is a salient but not inevitably predictive cue of their comprehension, they need to be supported in correctly interpreting or neglecting this cue. Actually, situation-model-approach interventions may decrease learners' reliance on effort by providing more valid cues. This assumption is supported by Schleinschok et al. (2017) who investigated the impact of generating a drawing after reading each passage on relative metacomprehension accuracy. Their results showed that, in general, the higher the learners' cognitive load during the learning task, the lower they judged their comprehension to be (cue utilization). However, compared with learners' judgments, their cognitive load was of limited predictive validity for their actual comprehension (cue diagnosticity). Therefore, cognitive load was used as a cue for judging comprehension but was hardly diagnostic for later comprehension performance. Yet, the learners in the drawing condition also used the quality of their drawings as a cue (cue utilization), which was highly predictive of their actual comprehension (cue diagnosticity), presumably contributing their higher level of accuracy (Experiment 1). Similarly, other cue-generation interventions might provide situation-model cues that learners can use in place of or in addition to their experienced mental effort, thereby enhancing relative metacomprehension accuracy. Moreover, the cue-selection intervention of setting the right test expectancy might alert learners about valid cues, decreasing their reliance on mental effort.

Limitations and Future Research

There are a few limitations that need to be considered when interpreting the present results and some important directions for future research. First, this meta-analysis reveals only that the level of accuracy changes as a function of situation-model-approach interventions. It leaves open the question of how the accuracy comes about. More concretely, although some studies investigated the (meta)cognitive processes underlying learners' judgments, such as their cue use (e.g., Thiede et al. 2010, Experiment 1), the precise mechanisms are not yet clear and alternative or additional processes, such as cognitive load, may play a role. Future research should therefore shed light on the proposed relations between cognitive load, cue use, and judgment accuracy, for example, by using self-reports of cue use or the think-aloud methodology during reading and judgment making. Based on this understanding, further interventions might be designed to optimize cognitive load and help learners focus on valid cues. Relatedly, different intervention types seem to support relative metacomprehension accuracy via different pathways, namely via the generation of, attention to, or selection of situation-model cues (Griffin et al. 2008, Experiment 2; Griffin et al. 2019b, Experiment 4; Wiley et al. 2008,

Experiment 2). Additionally, encoding-based and retrieval-based cue-generation interventions can be distinguished (Thiede et al. 2019). However, although the retrieval-based cue-generation interventions appear to be less beneficial for supporting comprehension than the encoding-based cue-generation interventions, there were no differences in relative metacomprehension accuracy between these two intervention types. Therefore, it does not seem to matter for relative metacomprehension accuracy when or through which cognitive processes situation-model cues are generated but rather that they are generated so that learners can use them to judge their comprehension. Yet, to more reliably reveal the impact of the kind of processing, that is, encoding versus retrieval, future studies should examine one particular intervention and only vary the processing task. For example, it could be varied whether learners draw directly during or after reading (cf. Thiede et al. 2019).

Second, there were only a small number of studies for some of the interventions under investigation, in particular, for delayed-diagram completion, self-explaining, and concept mapping. Therefore, the conclusions regarding these interventions should be treated with caution and further studies need to be conducted. Furthermore, it has to be emphasized that the studies on an intervention are not exact replications but differ in specific aspects concerning sample, design, materials, and procedure. However, we set several inclusion criteria, such as that only studies that examined expository texts and inference or application questions were eligible for inclusion. In addition, the studies were required to use a highly similar implementation of the respective intervention and a control condition in which the participants did not receive an intervention. In this way, we aimed at keeping a clear focus and at obtaining valid results concerning the investigated interventions. Yet, a larger empirical base would increase confidence in the results and allow for a more comprehensive examination of potential moderators, such as individual-difference variables (e.g., working-memory capacity, reading skill) or methodological aspects (e.g., text length, sample questions). In fact, the heterogeneity statistics for some of the interventions and intervention types indicated true between-studies variance that moderators might explain. Moreover, it should be noted that further interventions to promote the use of situation-model cues have been explored. For instance, the generation of drawings (Schleinschok et al. 2017; Thiede et al. 2019) or explanations (Fukaya 2013) also proved successful for enhancing relative metacomprehension accuracy. However, because there were fewer than two studies on these interventions that met the criteria for inclusion, they could not be investigated. Therefore, future research should replicate the studies on these interventions as well.

Third, the present study focused on metacomprehension accuracy in terms of relative accuracy. Absolute accuracy and bias are further measures of metacomprehension accuracy referring to the difference between judgments and performance. Specifically, absolute accuracy is the absolute difference and bias the signed difference reflecting over- and underconfidence. These two measures are also highly important in the context of self-regulated learning (see, e.g., De Bruin and Van Gog 2012; Dunlosky and Thiede 2013). For example, overconfidence can lead to the premature termination of studying and thus impair learning (e.g., Dunlosky and Rawson 2012). Compared with relative accuracy, absolute accuracy and bias assess different aspects of metacomprehension accuracy, have different strengths and limitations, and are often affected by different factors (e.g., Maki et al. 2005; Schraw 2009). However, to date, the impact of situation-model-approach interventions on absolute accuracy or bias has comparatively seldom been investigated. Consequently, future research should explore whether the interventions improve all measures, which might support self-regulated learning most effectively. Besides other measures, also other learning tasks might be considered. This study particularly focused on self-regulated learning from texts. A

specificity of learning in this context is the goal that learners build and monitor their situation model. Hence, it is questionable to what extent the present findings are transferable to other tasks, such as the learning of vocabularies or problem-solving, which aim at memorization or the acquisition of procedures. Although the situation-model approach gave rise to the interventions under investigation and helps explain their effects, for other learning tasks, other theoretical frameworks and interventions might be more useful. Thus, future research should assess what interventions are effective to support self-regulated learning involving other tasks.

Fourth, the interventions were less beneficial for improving comprehension than metacomprehension. This result indicates that conditions that lead to more accurate judgments and hence benefit learning in the long run do not necessarily benefit learning in the short run. It should be noted, however, that most studies specifically aimed at promoting metacomprehension, whereas it was not their goal to directly improve comprehension. For example, Jaeger and Wiley (2014, Experiment 2) remarked that their study on self-explaining did not involve practice or feedback, which has been central to self-explanation interventions designed to enhance comprehension (e.g., McNamara 2004), because their primary intention was to improve metacomprehension. In fact, for situation-model-approach interventions to increase relative metacomprehension accuracy, it is not necessary that learners construct a high-quality situation model. Instead, the critical feature is that learners use cues related to their situation model even if these cues indicate that their comprehension is poor. Nonetheless, in the ideal case, an intervention should improve all aspects of self-regulated learning, namely comprehension, metacomprehension accuracy, and regulation. Hence, further research should seek to uncover interventions meeting this requirement.

Conclusion

Interventions aimed at promoting learners' use of situation-model cues when judging their text comprehension can produce sizeable improvements in relative metacomprehension accuracy. On a practical level, instructors should apply these interventions to support learners in self-regulated learning from texts. Setting a comprehension-test expectancy seems to be particularly effective. That learners' tendency to rely on heuristic or memory-based cues can be overridden by a short statement informing them about the inferential nature of the test questions and a few sample test questions is promising. However, that such a minimal intervention can have such a large impact illustrates the ambiguity in what learners think it means to comprehend a text and suggests caution because their expectancy might easily be reverted. It seems appropriate to recommend giving learners a clear understanding of what it means to comprehend expository texts continually and from early on. Furthermore, there might not be a one-size-fits-all intervention for improving relative metacomprehension accuracy. Which particular intervention is most appropriate also depends on practical considerations, the specific situation, and the learners. Moreover, to achieve the maximum benefit, cue-generation, cue-attention, and cue-selection interventions might be combined, as they appear to support relative metacomprehension accuracy via different pathways (Griffin et al. 2008, Experiment 2; Griffin et al. 2019b, Experiment 4; Wiley et al. 2008, Experiment 2). However, further research is needed to clarify moderating effects and the mechanisms behind the situation-model-approach interventions.

On a theoretical level, there is the need for a comprehensive account integrating relevant theories, such as the cue-utilization framework (Griffin et al. 2009; cf. Koriat 1997), the situation-model approach (Griffin et al. 2019a; Wiley et al. 2016b), and the cognitive load

theory (Sweller 1988). The effort monitoring and regulation framework (De Bruin et al. 2020) represents a promising recent attempt in this direction. Such an account will provide the most complete explanation of why learners' metacomprehension is often inaccurate and how learners can be equipped with the metacognitive skills needed to effectively learn from reading texts.

Acknowledgments We thank all the researchers who contributed data to this meta-analysis and, in particular, those researchers who made efforts to provide requested information.

Funding Open Access funding enabled and organized by Projekt DEAL.

Compliance with Ethical Standards

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

References marked with an asterisk indicate studies included in the meta-analysis

- *Anderson, M. C. M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica*, *128*(1), 110–118. <https://doi.org/10.1016/j.actpsy.2007.10.006>.
- Baars, M., Leopold, C., & Paas, F. (2018). Self-explaining steps in problem-solving tasks to improve self-regulation in secondary education. *Journal of Educational Psychology*, *110*(4), 578–595. <https://doi.org/10.1037/edu0000223>.
- Baars, M., Vink, S., Van Gog, T., De Bruin, A. B. H., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction*, *33*, 92–107. <https://doi.org/10.1016/j.learninstruc.2014.04.004>.
- Baars, M., Wijnia, L., De Bruin, A. B. H., & Paas, F. (2020). The relation between students' effort and monitoring judgments during learning: A meta-analysis. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-020-09569-3>.
- Baker, J. M. C., & Dunlosky, J. (2006). Does momentary accessibility influence metacomprehension judgments? The influence of study–judgment lags on accessibility effects. *Psychonomic Bulletin & Review*, *13*(1), 60–65. <https://doi.org/10.3758/BF03193813>.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2013). *Comprehensive meta-analysis (Version 3)* [Computer software]. Biostat.
- Bugg, J. M., & McDaniel, M. A. (2012). Selective benefits of question self-generation and answering for remembering expository text. *Journal of Educational Psychology*, *104*(4), 922–931. <https://doi.org/10.1037/a0028661>.

- Carpenter, S., Endres, T., & Hui, L. (2020). Students' use of retrieval in self-regulated learning: Implications for monitoring and regulating effortful learning experiences. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-020-09562-w>.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Erlbaum.
- *Chiang, E. S., Theriault, D. J., & Franks, B. A. (2010). Individual differences in relative metacomprehension accuracy: Variation within and across task manipulations. *Metacognition and Learning*, 5(2), 121–135. <https://doi.org/10.1007/s11409-009-9052-6>.
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10(1), 101–129. <https://doi.org/10.2307/3001666>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- De Bruin, A. B. H., Roelle, J., Baars, M., & EFG-MRE. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda [Editorial]. *Educational Psychology Review*.
- *De Bruin, A. B. H., Thiede, K. W., Camp, G., & Redford, J. S. (2011). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology*, 109(3), 294–310. <https://doi.org/10.1016/j.jecp.2011.02.005>.
- De Bruin, A. B. H., & Van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, 22(4), 245–252. <https://doi.org/10.1016/j.learninstruc.2012.01.003>.
- Dunlosky, J., Baker, J. M. C., Rawson, K. A., & Hertzog, C. (2006). Does aging influence people's metacomprehension? Effects of processing ease on judgments of text learning. *Psychology and Aging*, 21(2), 390–400. <https://doi.org/10.1037/0882-7974.21.2.390>.
- Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16(4), 228–232. <https://doi.org/10.1111/j.1467-8721.2007.00509.x>.
- Dunlosky, J., & Rawson, K. A. (2005). Why does rereading improve metacomprehension accuracy? Evaluating the levels-of-disruption hypothesis for the rereading effect. *Discourse Processes*, 40(1), 37–55. https://doi.org/10.1207/s15326950dp4001_2.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction*, 22(4), 271–280. <https://doi.org/10.1016/j.learninstruc.2011.08.003>.
- Dunlosky, J., Rawson, K. A., & Hacker, D. J. (2002). Metacomprehension of science text: Investigating the levels-of-disruption hypothesis. In J. Otero, J. A. León, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 255–279). Erlbaum.
- Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, 24, 58–61. <https://doi.org/10.1016/j.learninstruc.2012.05.002>.
- Duval, S., & Tweedie, R. (2000a). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89–98. <https://doi.org/10.1080/01621459.2000.10473905>.
- Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.
- *Engelen, J. A. A., Camp, G., Van de Pol, J., & De Bruin, A. B. H. (2018). Teachers' monitoring of students' text comprehension: Can students' keywords and summaries improve teachers' judgment accuracy? *Metacognition and Learning*, 13(3), 287–307. <https://doi.org/10.1007/s11409-018-9187-4>.
- Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning*, 8(1), 1–18. <https://doi.org/10.1007/s11409-012-9093-0>.
- Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4), 702–718. <https://doi.org/10.1037/0278-7393.11.1.4.702>.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, 116(2), 119–136. <https://doi.org/10.1037/0096-3445.116.2.119>.
- Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, 37(7), 1001–1013. <https://doi.org/10.3758/MC.37.7.1001>.
- Griffin, T. D., Mielicki, M. K., & Wiley, J. (2019a). Improving students' metacomprehension accuracy. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (1st ed., pp. 619–646). Cambridge University Press.

- Griffin, T. D., Wiley, J., & Salas, C. R. (2013). Supporting effective self-regulated learning: The critical role of monitoring. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (Vol. 28, pp. 19–34). Springer.
- *Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, *36*(1), 93–103. <https://doi.org/10.3758/MC.36.1.93>.
- *Griffin, T. D., Wiley, J., & Thiede, K. W. (2019b). The effects of comprehension-test expectancies on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 1066–1092. <https://doi.org/10.1037/xlm0000634>.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. <https://doi.org/10.2307/1164588>.
- Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions* (Version 5.1.0). The Cochrane Collaboration. Retrieved from <https://handbook-5-1.cochrane.org>.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>.
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods*, *11*(2), 193–206. <https://doi.org/10.1037/1082-989X.11.2.193>.
- Hunter, J. E., & Schmidt, F. L. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Sage.
- Ikeda, K., & Kitagami, S. (2013). The interactive effect of working memory and text difficulty on metacomprehension accuracy. *Journal of Cognitive Psychology*, *25*(1), 94–106. <https://doi.org/10.1080/20445911.2012.748028>.
- *Jaeger, A. J., & Wiley, J. (2014). Do illustrations help or harm metacomprehension accuracy? *Learning and Instruction*, *34*, 58–73. <https://doi.org/10.1016/j.learninstruc.2014.08.002>.
- Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. Cambridge University Press.
- Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*(2), 133–159. [https://doi.org/10.1016/0749-596X\(90\)90069-C](https://doi.org/10.1016/0749-596X(90)90069-C).
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. <https://doi.org/10.1037/0096-3445.126.4.349>.
- Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The memorizing effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology*, *102*(3), 265–279. <https://doi.org/10.1016/j.jecp.2008.10.005>.
- Lin, L.-M., & Zabrocky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, *23*(4), 345–391. <https://doi.org/10.1006/ceps.1998.0972>.
- Lin, L.-M., Zabrocky, K. M., & Moore, D. (1996). The relations among interest, self-assessed comprehension, and comprehension performance in young adults. *Reading Research and Instruction*, *36*(2), 127–139. <https://doi.org/10.1080/19388079709558233>.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.
- Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Erlbaum.
- Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 663–679. <https://doi.org/10.1037/0278-7393.10.4.663>.
- Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 609–616. <https://doi.org/10.1037/0278-7393.16.4.609>.
- Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, *97*(4), 723–731. <https://doi.org/10.1037/0022-0663.97.4.723>.
- *Margolin, S. J., & Snyder, N. (2018). It may not be that difficult the second time around: The effects of rereading on the comprehension and metacomprehension of negated text. *Journal of Research in Reading*, *41*(2), 392–402. <https://doi.org/10.1111/1467-9817.12114>.
- McNamara, D. S. (2004). SERT: self-explanation reading training. *Discourse Processes*, *38*(1), 1–30. https://doi.org/10.1207/s15326950dp3801_1.
- Millis, K. K., Simon, S., & Tenbroek, N. S. (1998). Resource allocation during the rereading of scientific texts. *Memory & Cognition*, *26*(2), 232–246. <https://doi.org/10.3758/BF03201136>.

- Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-020-09541-1>.
- Pao, L. S. (2014). *Effects of keyword generation and peer collaboration on metacomprehension accuracy in middle school students* (Doctoral dissertation, Columbia University). Retrieved from <https://academiccommons.columbia.edu/doi/10.7916/D8HX19TV>.
- Prinz, A., Golke, S., & Wittwer, J. (2020). How accurately can learners discriminate their comprehension of texts? A comprehensive meta-analysis on relative metacomprehension accuracy and influencing factors. *Educational Research Review*. <https://doi.org/10.1016/j.edurev.2020.100358>.
- *Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, 28(6), 1004–1010. <https://doi.org/10.3758/BF03209348>.
- Rawson, K. A., O'Neil, R., & Dunlosky, J. (2011). Accurate monitoring leads to effective control and greater learning of patient education materials. *Journal of Experimental Psychology: Applied*, 17(3), 288–302. <https://doi.org/10.1037/a0024749>.
- *Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction*, 22(4), 262–270. <https://doi.org/10.1016/j.learninstruc.2011.10.007>.
- Roebers, C. M., Linden, N., & Howie, P. (2007). Favourable and unfavourable conditions for children's confidence judgments. *British Journal of Developmental Psychology*, 25(1), 109–134. <https://doi.org/10.1348/026151006X104392>.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Sage.
- Schleinschok, K., Eitel, A., & Scheiter, K. (2017). Do drawing tasks improve monitoring and control during learning from text? *Learning and Instruction*, 51, 10–25. <https://doi.org/10.1016/j.learninstruc.2017.02.002>.
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4(1), 33–45. <https://doi.org/10.1007/s11409-008-9031-3>.
- Shiu, L., & Chen, Q. (2013). Self and external monitoring of reading comprehension. *Journal of Educational Psychology*, 105(1), 78–88. <https://doi.org/10.1037/a0029378>.
- Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5–14. <https://doi.org/10.3102/0013189X08314117>.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4.
- *Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, 28(2), 129–160. [https://doi.org/10.1016/S0361-476X\(02\)00011-5](https://doi.org/10.1016/S0361-476X(02)00011-5).
- *Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95(1), 66–73. <https://doi.org/10.1037/0022-0663.95.1.66>.
- Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1267–1280. <https://doi.org/10.1037/0278-7393.31.6.1267>.
- *Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, 47(4), 331–362. <https://doi.org/10.1080/01638530902959927>.
- Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85–106). Routledge.
- Thiede, K. W., Redford, J. S., Wiley, J., & Griffin, T. D. (2012). Elementary school experience with comprehension testing may influence metacomprehension accuracy among seventh and eighth graders. *Journal of Educational Psychology*, 104(3), 554–564. <https://doi.org/10.1037/a0028660>.
- *Thiede, K. W., Redford, J. S., Wiley, J., & Griffin, T. D. (2017). How restudy decisions affect overall comprehension for seventh-grade students. *British Journal of Educational Psychology*, 87(4), 590–605. <https://doi.org/10.1111/bjep.12166>.
- Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology*, 81(2), 264–273. <https://doi.org/10.1348/135910710X510494>.
- Thiede, K. W., Wright, K. L., Hagenah, S., & Wenner, J. (2019). Drawings as diagnostic cues for metacomprehension judgment. In N. Feza (Ed.), *Metacognition in learning*. IntechOpen. Retrieved from

<https://www.intechopen.com/books/metacognition-in-learning/drawings-as-diagnostic-cues-for-metacomprehension-judgment>.

- Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, *35*(4), 668–678. <https://doi.org/10.3758/BF03193305>.
- *Van de Pol, J., De Bruin, A. B. H., Van Loon, M. H., & Van Gog, T. (2019). Students' and teachers' monitoring and regulation of students' text comprehension: Effects of comprehension cue availability. *Contemporary Educational Psychology*, *56*, 236–249. <https://doi.org/10.1016/j.cedpsych.2019.02.001>.
- Van de Pol, J., Van Loon, M. H., Van Gog, T., Braumann, S., & De Bruin, A. B. H. (2020). Mapping and drawing to improve students' and teachers' monitoring and regulation of students' learning from text: Current findings and future directions. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-020-09560-y>.
- Van Gog, T., Hoogerheide, V., & Van Harsel, M. (2020). The role of mental effort in fostering self-regulated learning with problem-solving tasks. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-020-09544-y>.
- *Van Loon, M. H., De Bruin, A. B. H., Van Gog, T., Van Merriënboer, J. J. G., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychologica*, *151*, 143–154. <https://doi.org/10.1016/j.actpsy.2014.06.007>.
- *Waldeyer, J., & Roelle, J. (2020). The keyword effect: A conceptual replication, effects on bias, and an optimization. *Metacognition and Learning*. <https://doi.org/10.1007/s11409-020-09235-7>.
- Weaver, C. A., Bryant, D. S., & Burns, K. D. (1995). Comprehension monitoring: Extensions of the Kintsch and Van Dijk model. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 177–193). Erlbaum.
- Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook on research in teaching* (3rd ed., pp. 315–327). Macmillan.
- Wiley, J., Griffin, T. D., Jaeger, A. J., Jarosz, A. F., Cushen, P. J., & Thiede, K. W. (2016a). Improving metacomprehension accuracy in an undergraduate course context. *Journal of Experimental Psychology: Applied*, *22*(4), 393–405. <https://doi.org/10.1037/xap0000096>.
- Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology*, *132*(4), 408–428. <https://doi.org/10.3200/GENP.132.4.408-428>.
- Wiley, J., Griffin, T. D., & Thiede, K. W. (2008). To understand your understanding one must understand what understanding means. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 817–822). Cognitive Science Society <https://pdfs.semanticscholar.org/6292/53f86294a75b66b6e9f296fc582c8cbbc75f.pdf>.
- Wiley, J., Griffin, T. D., & Thiede, K. W. (2016b). Improving metacomprehension with the situation-model approach. In K. Mokhtari (Ed.), *Improving reading comprehension through metacognitive reading strategies instruction* (pp. 93–110). Rowman & Littlefield.
- Wiley, J., Jaeger, A. J., Taylor, A. R., & Griffin, T. D. (2018). When analogies harm: The effects of analogies on metacomprehension. *Learning and Instruction*, *55*, 113–123. <https://doi.org/10.1016/j.learninstruc.2017.10.001>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.