Metacognition and Learning (2025) 20:17
https://doi.org/10.1007/s11409-025-09419-z

To what extent can think sheets with or without reflection prompts support self-regulated learning from text?

Anja Prinz-Weiß[1] · **Elisa Röhm**[2]

Received: 17 April 2024 / Accepted: 24 March 2025

Abstract
Misconceptions impair learners' text comprehension and judgment accuracy, which can hinder effective regulation processes. Think sheets are worksheets that require learners to contrast misconceptions with the correct information from a text. This study investigated to what extent completing a think sheet with ($n=30$) or without ($n=32$) reflection prompts compared to solely reading ($n=30$) benefits university students' text comprehension, transfer, judgment accuracy of comprehension and transfer, and adequacy of regulation decisions. Overall, the results showed that completing a think sheet provided little benefit compared to having students read about the misconceptions and correct information, and that there were no significant differences between completing a think sheet with and without reflection prompts. Nonetheless, an unanticipated advantage of completing a think sheet was observed concerning comprehension and transfer: When engaging in reading only, unexpectedly, students with fewer misconceptions before studying achieved poorer comprehension and transfer performance than students with more misconceptions. In contrast, when completing a think sheet, this influence of misconceptions did not occur. Judgment accuracy of comprehension and regulation decision adequacy did not depend on how many misconceptions students had and/or on the instructional method they received. Regarding judgment accuracy of transfer, students with more misconceptions before studying more strongly underestimated their ability to transfer their acquired knowledge, while there was no effect of the instructional method.

Keywords Judgment accuracy · Misconceptions · Reflection prompts · Regulation decisions · Text comprehension · Think sheets

Self-regulated learning from text requires students not only to engage in processes that directly promote understanding but also to accurately judge and regulate their own learning,

✉ Anja Prinz-Weiß
anja.prinz-weiss@ph-karlsruhe.de

[1] Department of Psychology, University of Education Karlsruhe, Erzbergerstr. 119, 76133 Karlsruhe, Germany

[2] Ludwig Maximilian University of Munich, Munich, Germany

Published online: 24 April 2025

such as by restudying poorly understood content. If students do not accurately judge their learning, this can lead to inappropriate regulation decisions (e.g., no restudying although it would be pertinent), impairing further knowledge acquisition (e.g., Rawson et al., 2011; Thiede et al., 2003). In fact, this intricate relationship between judgments of learning and regulation processes is central to many models of self-regulated learning (e.g., Nelson & Narens, 1990; Zimmerman, 2000).

Misconceptions, which are incorrect beliefs that contradict the current state of scientifically accepted information, can hinder self-regulated learning from text. In particular, it has been shown that students who have more misconceptions not only tend to achieve poorer comprehension and transfer performance but also to provide inaccurate judgments of their comprehension and transfer performance (e.g., Prinz et al., 2018, 2019). Therefore, it is highly relevant to address misconceptions in instruction. A method to do so are think sheets.

A think sheet is a worksheet that contains central questions about a topic as well as misconceptions concerning these questions. Learners have to provide the correct answers to the questions on the think sheet with the help of a text that provides the correct information (e.g., Dole & Smith, 1989; Prinz et al., 2021). Initial evidence indicates that completing a think sheet can counteract the detrimental influences of misconceptions on performance and judgment accuracy (Prinz et al., 2021). However, it is unclear whether think sheets also benefit learners' regulation decisions following their judgments. Moreover, it is an open question whether the effects of think sheets can be enhanced when learners are explicitly required to contrast the misconceptions and correct information. For example, to engage learners in active comparison, reflection prompts, which are scaffolds to guide learners' attention towards specific aspects of the learning content, could be used. Thus, in the present study, we examined to what extent completing a think sheet with or without reflection prompts compared to solely reading would support university students' comprehension, transfer, judgment accuracy of comprehension and transfer, and adequacy of regulation decisions.

The detrimental effects of misconceptions

Misconceptions occur in many domains and about many topics. One domain in which misconceptions are widespread is statistics. For instance, misconceptions exist with regard to the statistical concept of covariance (e.g., Huck, 2016). As examples, many students mistakenly assume that covariance proves a causal relationship between two variables or that covariance is a standardized measure that takes on values within a finite range (e.g., Prinz et al., 2018, 2019, 2021). Statistical misconceptions often have a coherent internal structure, enabling the continuous derivation of explanations without realizing that they are incorrect. For example, learners who have the misconception that covariance proves causality might consistently misinterpret the relationship between two variables by assuming that one variable causally affects the other variable. This feature typically makes the revision of the misconceptions, that is, conceptual change, hard to achieve (e.g., Lewandowsky et al., 2012).

Furthermore, research has shown that statistical misconceptions can have detrimental effects on performance and judgment accuracy. Specifically, university students with a higher number of misconceptions about covariance not only achieved poorer text comprehension, but also provided more overconfident judgments of their comprehension compared to students with fewer misconceptions (Prinz et al., 2018, 2019, 2021). A similar influ-

ence of the misconceptions was also partially found on transfer and judgment accuracy of transfer (Prinz et al., 2018, 2019; not in Prinz et al., 2021). Presumably, students who have misconceptions incorporate their false beliefs into the inferences they make while reading a text, which impairs their comprehension and application of it (e.g., Kendeou & Van den Broek, 2005; see also Prinz et al., 2019, 2021). Moreover, when making judgments, the students might rely on the amount of information they can access from memory, without recognizing that this information is at least partly incorrect, which leads to overconfident comprehension judgments (e.g., Dunlosky et al., 2005; see also Prinz et al., 2019, 2021).

Think sheets as a method to address misconceptions

Many methods have been developed to improve students' reading comprehension. For example, successful interventions include questioning the text, predicting contents, and summarizing main ideas (e.g., McNamara, D. S. 2007). However, as has recently been emphasized, it is highly important that the applied interventions align with the readers' characteristics (e.g., McMaster & Kendeou, 2023). If readers have misconceptions, this needs to be taken into account. An effective intervention to do so are refutation texts, which are texts that explicitly state misconceptions before refuting them with the correct information (e.g., Danielson et al., 2024). Apart from refutation texts, interventions that consider learners' misconceptions when acquiring knowledge from reading have been scarcely investigated. Think sheets represent such a potential alternative intervention. A think sheet is a worksheet on which central questions about a topic and misconceptions concerning these questions are listed or entered by the learners. In addition to the think sheet, a text that provides the correct information is given to the learners. Their task is to extract the correct information from the text and to use it to write down the correct answers to the questions on the think sheet (Dole & Smith, 1989; see also Brummer & Macceca, 2013; Fang et al., 2010).

So far, the effectiveness of think sheets appears to have been investigated in only two studies. Dole and Smith (1989) conducted the first study with fifth-grade students in which one group was taught a physics unit with standard text-based instruction, whereas another group worked with think sheets. In their lessons, students in the think-sheet group were first asked to write down their own preexistent beliefs about questions in a column on a think sheet. Then, while reading a physics text, they recorded the correct answers to the questions based on the textual information in the next column. The think sheets contained additional columns where the students had to check if (a) the text ideas were the same as their preexistent ideas, (b) the text ideas added information to their preexistent ideas, (c) the text ideas conflicted with their preexistent ideas, and (d) whether the text was confusing. The results showed that the students who worked with think sheets acquired a more scientifically correct comprehension than the students who learned with the standard instructions.

Prinz et al. (2021) followed up on this study and examined to what extent completing a think sheet, reading a refutation text, or engaging in standard reading affected university students' performance and judgment accuracy. Specifically, students in the think-sheet group received a think sheet containing central questions and related misconceptions about covariance as well as a text providing the correct information. Their task was to complete the think sheet by answering the questions with the correct information from the text. Students in the refutation-text group received a refutation text to read. For each misconception about covariance, this text first explicitly stated the misconception, then warned about the fact that

the misconception is incorrect, and then provided the correct information. Students in the control group received a standard text providing the correct information about covariance as well as the central questions and misconceptions to read. The results showed that both methods, completing the think sheet and reading the refutation text, counteracted the detrimental impact of pre-existing misconceptions on students' comprehension and overconfidence when predicting their comprehension occurring in the control group. Moreover, completing the think sheet helped students who had more misconceptions in the beginning to transfer their newly acquired comprehension, although they underestimated their ability to do so. In line with the Knowledge Revision Components (KReC) framework (Kendeou & O'Brien, 2014), the two methods might have facilitated coactivation and further conceptual-change processes. More precisely, the KReC framework (Kendeou & O'Brien, 2014) is a theoretical model that suggests that the central prerequisite for knowledge revision is that the misconception and correct information are simultaneously activated in a learner's memory (coactivation principle). Once coactivation has occurred, the new correct information can be integrated into the existing knowledge structures, whereby the misconception should ideally be transformed and outdated (integration principle). Lastly, as an increasing amount of correct information becomes integrated in a learner's knowledge structures, it begins to dominate the overall network (competing activation principle). When completing a think sheet in particular, learners are required to extract the correct information from a text and to set it against the respective misconceptions. In this way, the think sheet in the study by Prinz et al. (2021) likely evoked coactivation and additional conceptual-change processes, resulting in an enhanced comprehension and more careful evaluation of one's comprehension. Moreover, completing a think sheet is a generative task, because learners actively have to make sense of the material by filtering the relevant information from a text, organizing the information mentally, and integrating it into their existing knowledge structures. Generative tasks promote deep and transferable learning (e.g., Fiorella & Mayer, 2016). Accordingly, completing the think sheet might have provided additional benefits for transfer in the study by Prinz et al. (2021).

To conclude, the two previous studies suggest that think sheets might be an effective method when it comes to addressing misconceptions. However, some aspects remain unclear. First, it is unclear whether think sheets cannot only support the accuracy with which students judge their comprehension but also the adequacy of their regulation decisions such as about whether they should engage in restudying or not. Ideally, learners should base their regulation decisions on their judgments of learning. For example, when learners perceive that they have not yet comprehended the contents well, they should decide to restudy. In contrast, when learners perceive that they have already comprehended the contents well, decisions not to restudy seem adequate. Accordingly, the adequacy of regulation decisions is typically operationalized via the extent to which learners' restudy decisions match with their judgments (regardless of the accuracy of those judgments, because learners have no access to their actual level of understanding; e.g., van de Pol et al., 2020, 2021). Second, in the study by Dole and Smith (1989), the think sheets contained additional columns that required the students to contrast their preexistent ideas and the information from the text. However, it is unclear to what extent such reflection prompts can further enhance the effectiveness of think sheets in supporting self-regulated learning from text.

Reflection prompts as a means to foster active comparison

Instructional prompts are scaffolds to stimulate certain activities in learners. For example, they might induce the recall of information or the execution of certain processes (e.g., Bannert, 2009). Prompts can be designed and implemented in various ways (e.g., via hints, explicit instructions, or questions), but have the common goal to direct learners' attention to specific aspects of the learning process (e.g., Wirth, 2009). Metacognitive prompts are supposed to induce activities such as planning, monitoring, controlling, evaluating, or reflecting (e.g., Bannert, 2009). One type of metacognitive prompts are reflection prompts, which focus learners' attention on their own understanding or used activities (e.g., Bannert, 2006; Lin & Lehman, 1999). Several studies have shown that knowledge acquisition and transfer can be enhanced by metacognitive prompts (e.g., Bannert et al., 2015; Bannert & Reimann, 2012; Johnson et al., 2011; Kramarski & Michalsky, 2013) and reflection prompts in particular (e.g., Bannert, 2006; Davis & Linn, 2000; Lin & Lehman, 1999; Stark & Krause, 2009). For example, Davis and Linn (2000) examined whether reflection prompts improve eighth-grade students' understanding of thermodynamics. Specifically, the students received prompts to plan ahead (e.g., prompts to think about the activity on which they were embarking) and to reflect (e.g., prompts to think about what evidence or claims they did not understand well). The results showed that the prompts helped the students to integrate their knowledge and develop a coherent understanding. Stark and Krause (2009) examined the effects of reflection prompts in a statistics e-learning environment. In the environment, students had to complete tasks and decide whether they wanted to receive support in the form of worked-out examples (a) immediately or (b) after solving the tasks. Students who received reflection prompts were required to provide short justifications for their decisions. The results showed that prompted students performed better both immediately after learning and on a follow-up test compared to students who received no prompts.

More recently, a few studies applied reflection prompts as a means to facilitate misconception revision (Bascandziev, 2020, 2024; Theobald et al., 2024). Bascandziev (2020, 2024) showed that adults were more successful in correcting their physics misconceptions through a training with thought experiments when they were prompted to think about the implications of their responses and how they relate to their previous beliefs. In the study by Theobald et al. (2024), children had to make predictions when learning about the principles of water displacement. One group of children additionally received reflection prompts that required them to think about how their predictions relate to what they had learned. The results revealed that children in the prediction-with-reflection-prompts group were better in recognizing incorrect predictions, more likely to make correct predictions after recognizing incorrect predictions, and faster to revise misconceptions and make correct predictions during learning. After learning, however, no effects of the prompts on posttest and transfer test performance were found, potentially because the effects of generating predictions were already quite strong. Theobald et al. (2024) suggested that the reflection prompts might have helped revise misconceptions during learning indirectly through the improved recognition of conflict and directly by supporting knowledge integration processes. Similarly, reflection prompts within think sheets might be useful to aid learners' recognition of conflicting information on the one hand and knowledge integration on the other hand. However, it is unclear whether think sheets that contain reflection prompts can facilitate the revision of misconceptions and support performance and judgment accuracy even more effectively.

Present study

Think sheets are a promising method worth further investigation due to several reasons. On a theoretical level, with the exploration of think sheets, the principles of the KReC framework (Kendeou & O'Brien, 2014) can be probed. Specifically, it can be explored to what extent think sheets with and without reflection prompts facilitate coactivation and knowledge integration and thereby self-regulated learning. On a practical level, compared to refutation texts, which currently seem to be the predominantly researched and used method to address misconceptions, think sheets can be applied without rewriting instructional texts. In addition, to keep learners motivated, it is beneficial to alternate instructional methods (e.g., Smith & Ragan, 2005).

Hence, the first aim of this study was to examine to what extent completing a think sheet would counteract detrimental influences of misconceptions on comprehension, transfer, judgment accuracy of comprehension and transfer, and adequacy of regulation decisions. The second aim of this study was to explore whether these potential compensatory effects would differ depending on whether students complete a think sheet with or without reflection prompts. Due to their high prevalence among students (e.g., Prinz et al., 2018, 2019, 2021), we focused on misconceptions about the statistical concept of covariance.

The two prior studies conducted on think sheets revealed that by requiring students to contrast misconceptions with the correct information, think sheets can support comprehension and transfer (Dole & Smith, 1989; Prinz et al., 2021). Moreover, reflection prompts proved useful for revising misconceptions and supporting knowledge acquisition and transfer (e.g., Bannert, 2006; Stark & Krause, 2009; Theobald et al., 2024). However, the effects of reflection prompts have not yet been examined in the context of think sheets. Accordingly, we hypothesized that, under reading-only conditions, students with more misconceptions would achieve poorer comprehension and transfer performance, whereas these negative effects of misconceptions would not occur when completing a think sheet (*comprehension hypothesis* and *transfer hypothesis*). In addition, we explored whether the assumed compensatory effects would be more pronounced with a think sheet that contains reflection prompts than with a think sheet that does not contain reflection prompts.

Previous research indicated that completing a think sheet can also benefit the accuracy with which students judge their learning (Prinz et al., 2021). Moreover, reflection prompts are metacognitive prompts that aim at increasing monitoring and evaluation processes (e.g., Bannert, 2006, 2009). However, their effects on judgment accuracy have not yet been investigated. Therefore, we assumed that, under reading-only conditions, students with more misconceptions would exhibit greater inaccuracy in their judgments of comprehension and transfer, but that these detrimental effects of misconceptions would not emerge when completing a think sheet (*comprehension-judgment hypothesis* and *transfer-judgment hypothesis*). We also explored whether the expected compensatory effects would be more pronounced with a think sheet that contains reflection prompts than with a think sheet that does not contain reflection prompts.

Finally, it has been shown that more accurate judgments often go along with more adequate regulation decisions (e.g., Thiede et al., 2003). Thus, we expected that, under reading-only conditions, students with more misconceptions would show less adequate restudy decisions, but that this negative impact of misconceptions would not occur when completing a think sheet (*regulation-adequacy hypothesis*). Again, we explored whether this compensa-

tory effect would be more pronounced with a think sheet that contains reflection prompts than with a think sheet that does not contain reflection prompts.

Method

Participants

To determine the minimum number of required participants, an a priori power analysis was conducted with G*Power (Faul et al., 2007). We used a power of 90% to reduce the chance that actual effects would not be detected. Based on previous research (e.g., Prinz et al., 2021), we expected to find medium effects of completing a think sheet in our study. The analysis revealed that a sample of 88 participants would be needed (power analysis for linear multiple regression: R^2 increase: $f^2 = 0.15$, $\alpha = 0.05$, power = 90%, number of tested predictors = 2, total number of predictors = 5). To account for potential drop out, we collected data from $N = 99$ university students.

Participants were required to study at a German university and to have attended at least one statistics course that covered the topic of covariance to ensure that they had some prior knowledge and possibly misconceptions. The data of $n = 7$ participants was excluded because they completed the study although they had not attended such a statistics course. Hence, the final sample consisted of $N = 92$ participants who were on average 22.29 ($SD = 3.64$) years old, with 58% being female and 42% being male (none non-binary). On average, the participants had studied for 4.26 ($SD = 2.27$) semesters and had attended 1.72 ($SD = 0.86$) statistics courses that covered covariance. Specifically, 42 participants (46%) had attended 0.5 or 1.0 course, 36 participants (39%) had attended 1.5 or 2.0 courses, 10 participants (11%) had attended 2.5 or 3.0 courses, and 4 participants (3%) had attended between 3.5 and 5.0 courses. As their major subject, the participants studied educational science (33%), psychology (20%), economics (20%), politics (10%), cultural studies (3%), philosophy (3%), informatics (3%), social sciences (2%), sociology (2%), a foreign language (2%), or medicine (1%). Fifty percent of the participants also studied a minor subject: cognitive science (24%), a foreign language (7%), economics (7%), politics (3%), cultural studies (2%), geography (2%), biology (1%), psychology (1%), material sciences (1%), music (1%), or medicine (1%).

Participants were recruited via social media, e-mail, and a post on the website of the institute where the study was conducted. They were informed about who the researchers were, the study's focus on self-regulated learning from text, and its estimated duration of 45 min. Participants received course credit (1 experimental hour) or financial reward (8 Euros) for taking part in the study.

Design

The study had an experimental between-subjects design with the participants being randomly assigned to the think-sheet group ($n = 32$), the think-sheet-reflection group ($n = 30$), or the control group ($n = 30$). In the online study environment, all participants were given a statistics text that provided the correct information about covariance. Besides, participants in the *think-sheet group* received four central questions about covariance and respective

misconceptions. Their task was to write down the correct answers to the four questions with the help of the text. Participants in the *think-sheet-reflection group* received the same material and task with the addition of reflection prompts. Specifically, after providing the correct answer to each question, the participants answered four reflection prompts. In addition to the statistics text, participants in the *control group* received the four questions with the respective misconceptions to read, but no additional task.

The number of misconceptions about covariance that participants initially held was measured and not experimentally manipulated. The dependent variables were comprehension, transfer, judgment accuracy of comprehension, judgment accuracy of transfer, and adequacy of regulation decisions.

Materials

Statistics text

The statistics text on covariance was adapted from a widely used German statistics textbook (Bortz & Schuster, 2010; cf. also Prinz et al., 2018, 2019, 2021). The text described four aspects of covariance that were tapped by the four central questions (i.e., the magnitude of covariance is related to the spread of the data points in a scatter plot, zero covariance implies the absence of a linear relationship, covariance indicates the extent to which two variables vary together, covariance is a nonstandardized statistic depending on the measuring units of the variables). In addition, the text explained how covariance is calculated. The text included 539 words (27 sentences) and had a Flesch-Reading-Ease score of 37 (Flesch, 1948), indicating that it was rather difficult to read but appropriate for the sample of university students.

Think sheet

The think-sheet group received the think sheet, which was implemented in the online study environment via four blocks (see Fig. 1 for an example). Each block started with a central question concerning covariance. Then, a typical misconception concerning the question was described. Finally, there was a blank space in which the participants were required to fill in

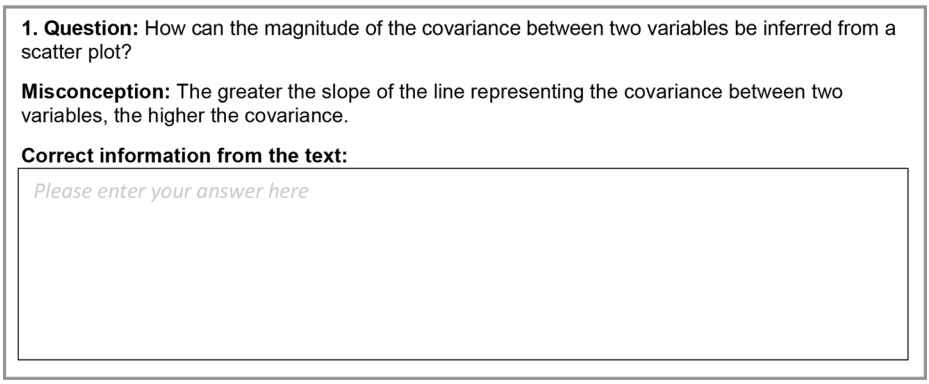


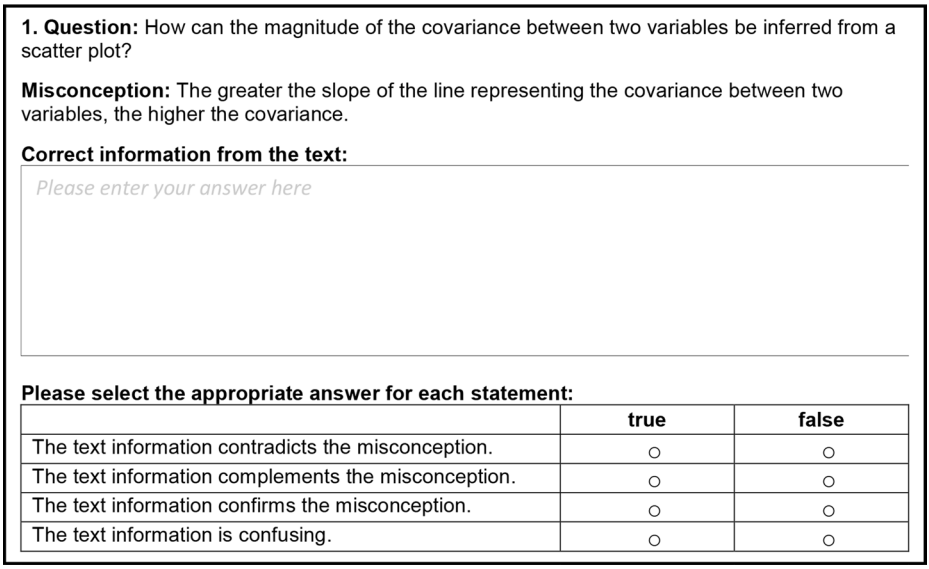
Fig. 1 Example block of the think sheet

the correct information concerning the question provided in the text. The four questions and related misconceptions were the following:

(1) Question: How can the magnitude of the covariance between two variables be inferred from a scatter plot?; Misconception: The greater the slope of the line representing the covariance between two variables, the higher the covariance.
(2) Question: What does a covariance of zero between two variables mean?; Misconception: A covariance of zero means that there is no relationship between the two variables at all.
(3) What does covariance indicate about the cause-and-effect relationship between two variables?; Misconception: Covariance between two variables indicates that one variable causally influences the other variable and thus that there is a causal relationship.
(4) Question: How can the statistical measure of covariance be interpreted?; Misconception: Like the correlation coefficient, covariance is a standardized measure that takes on values within a specific range (-1 to +1), allowing for direct comparisons of different covariance values.

Think sheet with reflection prompts

The think-sheet-reflection group received the think sheet with reflection prompts. Specifically, in addition to the think sheet described previously, a matrix with four reflection prompts was presented on each block (see Fig. 2 for an example). The prompts contained the following statements adapted from Dole and Smith (1989): the text information contradicts the misconception, the text information complements the misconception, the text information confirms the misconception, and the text information is confusing. Hence, after reading a question with the associated misconception and formulating an answer using the correct



1. Question: How can the magnitude of the covariance between two variables be inferred from a scatter plot?

Misconception: The greater the slope of the line representing the covariance between two variables, the higher the covariance.

Correct information from the text:

Please enter your answer here

Please select the appropriate answer for each statement:

	true	false
The text information contradicts the misconception.	○	○
The text information complements the misconception.	○	○
The text information confirms the misconception.	○	○
The text information is confusing.	○	○

Fig. 2 Example block of the think sheet with reflection prompts

textual information, the participants had to complete the prompts by indicating whether they thought each statement was true or false.

Control sheet

The control group received the control sheet, which contained the four central questions concerning covariance and the respective misconceptions that were also presented on the think sheet (see Fig. 3 for an example). This was done to keep the amount of information presented to all participants constant. The control sheet, however, did not include a blank space where the correct information from the text had to be entered or prompts.

Measures

Misconceptions before studying

The misconceptions participants held before studying were assessed with a misconceptions test. The test was adopted from previous research and consisted of four questions assessing the four common misconceptions about covariance that were later addressed in the learning materials (Prinz et al., 2018, 2019, 2021). As mentioned earlier, the four misconceptions were that (1) the magnitude of covariance is related to the slope of the fit line in a scatter plot, (2) zero covariance shows that there is no association between two variables at all, (3) covariance proves a causal relationship between two variables, and (4) covariance is a standardized statistic. The questions had a single-choice format with four response options. One option represented the misconception, one option represented the correct answer, and the two remaining options represented incorrect answers (see Fig. 4 for an example). Consequently, participants could record between 0 and 4 misconceptions. By adding two incorrect response options, we wanted to reduce the effect of guessing and provide further alternatives that, although not common misconceptions, could be naïve assumptions of students. Accordingly, the results showed that participants also selected these incorrect response options, but not with high frequency: 63 participants (69%) had no incorrect knowledge, 24 participants (26%) gave one incorrect answer, and 5 participants (5%) provided two incorrect answers. Furthermore, the instructional materials (i.e., the text and worksheets) were specifically designed to address the four misconceptions and not other incorrect assumptions. Thus, to be able to assess the specific effects of the misconceptions and the effectiveness of the materials in counteracting them, we did not classify all incorrect response options as misconceptions, but only the one actually representing the misconception. The four questions as well as the four response options were displayed in a randomized order.

1. Question: How can the magnitude of the covariance between two variables be inferred from a scatter plot?

Misconception: The greater the slope of the line representing the covariance between two variables, the higher the covariance.

Fig. 3 Example block of the control sheet

In a study, sports scientists from a university determined the covariance between the height and the time for a 100-m race of 20 sprinters. In his calculation, sports scientist A quantified time in seconds. When his colleague, sports scientist B, checked again, he quantified time in milliseconds. Which of the following statements about the covariances calculated by the two sports scientists is correct?

- ☐ The two sports scientists received the same covariance because it does not matter that sports scientist A used seconds and sports scientist B used milliseconds for the calculation. *(misconception)*
- ☐ The calculations of both sports scientists yielded no covariance because one cannot calculate covariance from time data. *(incorrect)*
- ☐ No statement about the two covariances can be made because it is unknown whether the time and height variables were linear. *(incorrect)*
- ☐ Sports scientist B obtained a higher covariance than sports scientist A because milliseconds yield bigger numbers than seconds. *(correct)*

Fig. 4 Sample question of the misconceptions test

Comprehension

To assess comprehension of the statistics text, the same four questions of the misconceptions test were used. The four questions and response options were again displayed in a randomized order to minimize memory effects. The participants received 1 point for each correct answer. Thus, they could achieve between 0 and 4 points.

Transfer

Based on previous research (Prinz et al., 2018, 2019, 2021), transfer was measured by assessing comprehension of an educational research report on scientific literacy with four questions. The research report was adapted from PISA (Organisation for Economic Co-operation and Development, 2006) and described findings involving the covariance between aspects of scientific literacy (e.g., familiarity with environmental issues) and academic achievement in science. Specifically, the educational research report presented information that learners with misconceptions about covariance were likely to misunderstand. The text comprised 663 words and had a Flesch-Reading-Ease score of 26 (Flesch, 1948). The four transfer questions addressed the same four aspects about covariance as the comprehension questions. However, the questions were modified and embedded in different cover stories to prevent memory and fatigue effects from repeated exposure to the same questions. The participants received 1 point for each correct answer. Thus, they could achieve between 0 and 4 points.

Judgment accuracy

As common in research on judgment accuracy in the context of self-regulated learning from text (e.g., Braumann et al., 2024; Thiede et al., 2012; see also Prinz et al., 2020), after studying, the participants predicted how many of the upcoming test questions (0 to 4) they would presumably answer correctly. They made their predictions separately for the comprehension and transfer questions. Beforehand, the participants were informed that the comprehension

questions would tap their understanding of the presented information, whereas the transfer questions would require them to apply their acquired understanding.

Judgment accuracy was operationalized as bias, which is the signed difference between a participant's judged and actual number of correct answers (e.g., Golke et al., 2022; Maki et al., 2005; Schraw, 2009). Thus, judgment accuracy ranged from −4.00 to +4.00. A positive value indicated overconfidence, a negative value underconfidence, and a value of zero a perfectly accurate judgment.

Regulation decision adequacy

In line with the majority of previous research conducted on regulation decisions in the context of self-regulated learning from text (e.g., Thiede et al., 2003, 2012; van de Pol et al., 2019), after providing their judgments, the participants indicated whether they would like to restudy the materials to improve their understanding ("yes" or "no").

Regulation decision adequacy was operationalized in absolute terms by relating participants' restudy decisions to their comprehension judgments based on the approach by van de Pol et al. (2020, 2021). Table 1 shows how the scores were allocated. Lower judgments (indicating that a learner perceives their understanding as limited) accompanied by decisions to restudy as well as higher judgments (indicating that a learner perceives their understanding as good) accompanied by decisions not to restudy resulted in higher regulation decision adequacy (i.e., higher values). In contrast, higher judgments (indicating that a learner perceives their understanding as good) accompanied by decisions to restudy as well as lower judgments (indicating that a learner perceives their understanding as limited) accompanied by decisions not to restudy resulted in lower regulation decision adequacy (i.e., lower values). Regulation decision adequacy ranged from 0 to 1, with higher values indicating greater regulation decision adequacy.

Procedure

Participants took part in the study online on their own computers at home during the COVID-19 pandemic in spring 2021. In the introduction of the study, the conditions of participation and general instructions were provided. Then, the participants completed the misconceptions test. Next, they received the statistics text and further learning mate-

Table 1 Coding of regulation decision adequacy

Restudy decision (0 = no, 1 = yes)	Comprehension judgment (0 to 4)	Regulation-decision-adequacy score
1	0	1.00
1	1	0.75
1	2	0.50
1	3	0.25
1	4	0.00
0	0	0.00
0	1	0.25
0	2	0.50
0	3	0.75
0	4	1.00

rial depending on their condition. Specifically, participants in the think-sheet group received the think sheet. They were asked to answer the central questions using the correct information from the statistics text. Participants in the think-sheet-reflection group were presented with the think sheet that additionally contained reflection prompts. They had to answer the central questions using the correct information from the statistics text and complete the prompts. Participants in the control group were provided with the control sheet. Their task was to read the materials. All participants were instructed to study the provided materials thoroughly and told that they would later be tested on the content. They had 15 min for studying. Thereafter, the participants made their judgments of comprehension and transfer. To do so, they were informed about the nature of the questions, the question format, and the number of questions. Then, the participants made their restudy decisions. They were told that, due to time constraints, they would not actually restudy. Following that, the participants answered the comprehension questions. Thereafter, they read the educational research report for 8 min and answered the transfer questions. Finally, the participants completed demographic questions. The study took approximately 45 min.

Analysis

To examine our hypotheses, we conducted multiple regression analyses with number of misconceptions as the predictor variable and instructional method as the multicategorical moderator variable following the approach by Hayes (2018). Indicator coding was used to generate two grouping variables (i.e., think sheet/control and think sheet reflection/control) coding the received instructional method, with the control group being the reference group. The predictor variable number of misconceptions was mean centered. Therefore, the regression coefficients for the grouping variables (i.e., the regression coefficients for the variables think sheet/control and think sheet reflection/control in the Tables estimate the difference in the dependent variable between the control group and the think-sheet or the think-sheet-reflection group at an average number of misconceptions (without mean centering, the regression coefficients would provide the effects when the number of misconceptions is zero). In each regression model, the number of misconceptions, the two grouping variables, and two interaction terms between the number of misconceptions and the grouping variables were entered. An overall interaction between the number of misconceptions and instructional method was indicated by a significant change in R^2 (i.e., ΔR^2) when the interaction terms were included in the regression model compared with when they were not. In this case, we further probed the conditional effects and individual interactions.

Results

Table 2 presents the descriptive statistics and intercorrelations among the study variables. Moreover, Table 3 shows the descriptive statistics for the three groups separately. For all analyses, an alpha level of 0.05 was used.

Table 2 Descriptive statistics and intercorrelations among the variables

Measure	1	2	3	4	5	6	7	8
1. No. of misconceptions before studying	–							
2. Comprehension	−.05	–						
3. Transfer	.07	.60**	–					
4. Judgment accuracy of comprehension	−.13	−.67**	−.17	–				
5. Judgment accuracy of transfer	−.24*	−.45**	−.77**	.40**	–			
6. Regulation decision adequacy	−.05	.27*	.38**	.07	−.22*	–		
7. No. of statistics courses on covariance	−.12	.02	.07	−.13	−.15	.00	–	
8. No. of semesters	.14	.06	.08	−.19	−.13	−.15	.35**	–
M	1.86	2.75	2.62	−0.05	−0.47	.63	1.72	4.26
SD	1.04	1.06	1.10	1.02	1.06	.24	0.86	2.27

*p<.05. **p<.001

Table 3 Means and standard deviations of the variables in the three groups

	Think-sheet group (n=32)		Think-sheet-reflection group (n=30)		Control group (n=30)	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
No. of misconceptions before studying	1.78	1.18	1.80	1.06	2.00	0.87
Comprehension	2.97	0.90	2.87	1.07	2.40	1.13
Transfer	2.75	0.95	2.67	1.16	2.43	1.19
Judgment accuracy of comprehension	-0.19	0.82	-0.10	1.03	0.13	1.96
Judgment accuracy of transfer	-0.53	1.05	-0.43	1.04	-0.43	1.14
Restudy decision adequacy	.65	.26	.63	.25	.60	.22
No. of statistics courses on covariance	1.75	0.80	1.42	0.76	1.98	0.94
No. of semesters	4.17	2.25	4.04	2.28	4.56	2.33

Preliminary analyses

Across groups, participants held 1.86 (SD=1.04) misconceptions before studying. Only 11% had no misconceptions, and 65% held at least two misconceptions. This shows that the misconceptions about covariance were quite prevalent among the participants.

There was a significant difference between the three groups concerning the number of attended statistics courses that covered covariance, $F(2, 89)=3.48$, $p=.035$, $\eta^2=.07$ (medium effect). However, it should be noted that in each of the three groups, the majority of the participants had attended between 1.0 and 2.0 courses: control group: 24 participants (80%), think-sheet group: 25 participants (78%), think-sheet-reflection group: 26 participants (87%). Hence, the somewhat higher average number of courses attended in the control and think-sheet groups compared to the think-sheet-reflection group likely stems from a few participants who had attended more courses (control group: 4 participants had attended between 3.0 and 5.0 courses, think-sheet group: 7 participants had attended 3.0 courses, think-sheet-reflection group: 2 participants had attended 3.0 and 4.0 courses). In addition, the number of misconceptions about covariance, as a more specific measure of prior knowledge, did not significantly differ between the groups, $F(2, 89)=0.41$, $p=.668$,

$\eta^2 = .01$ (small effect). The three groups did also not significantly differ from each other with regard to the number of semesters, $F(2, 79) = 0.37$, $p = .694$, $\eta^2 = .01$ (small effect). The number of attended statistics courses on covariance and the number of semesters did not significantly correlate with the number of misconceptions about covariance (predictor) or any of the dependent variables, all ps $> .05$. Therefore, they were not considered as covariates. Note that the results of the main analyses remained the same (although the exact numbers differed) when the number of attended statistics courses on covariance and the number of semesters were included as covariates.

Quality of completed think sheets

Answers to the central questions

As a treatment check, we examined the participants' answers to the four questions on the think sheet. First, the results showed that there were very few missing answers (i.e., two empty response fields and two answers indicating no knowledge out of a total of 248 participant responses) and no nonsensical answers. Hence, the participants put in effort to complete the think sheet. Second, an answer was assigned 1 point when it was correct *and* complete and 0 points when it was incorrect or incomplete. Hence, the participants could achieve a maximum of 4 points. The results showed that 30 of the 32 participants in the think-sheet group and 28 of the 30 participants in the think-sheet-reflection group scored at least two points (18 and 21 participants achieved the maximum score of 4 points, respectively). This outcome indicates that the participants completed the think sheet thoroughly. Note also that the results of the main analyses were the same when the four participants (two in each group) who scored below 2 points were excluded from the sample.

Furthermore, we examined whether the quality with which the think sheet was completed played a role for the effectiveness of the think sheet. To do so, we conducted moderation analyses on the outcome variables with the number of points on the think sheet as the predictor variable and instructional method (think-sheet group vs. think-sheet-reflection group) as the moderator variable. There were significant positive main effects of the number of points concerning comprehension, $b = 0.55$, $SE = 0.10$, $t(59) = 5.31$, $p < .001$, transfer, $b = 0.32$, $SE = 0.13$, $t(59) = 2.48$, $p = .016$, and regulation decision adequacy, $b = 0.08$, $SE = 0.03$, $t(59) = 2.50$, $p = .015$. Thus, when participants completed the think sheet more accurately (i.e., gave more complete and correct answers), their comprehension, transfer performance, and regulation decision adequacy were higher. Note that concerning judgment accuracy of comprehension, the negative main effect just failed to reach significance, $b = -0.23$, $SE = 0.12$, $t(59) = -2.00$, $p = .050$, and suggested that when participants completed the think sheet more accurately, they tended to be less overconfident in their comprehension. No significant interaction effects appeared, indicating that the effects did not differ between the think-sheet and think-sheet-reflection groups.

Answers to the reflection prompts

We further examined the participants' answers to the reflection prompts in the think-sheet-reflection group (see Table 4). Ideally, the participants should have indicated that the statement "the text information contradicts the misconception" is true and, conversely, that the

Table 4 Percentage of agreement (i.e., "True") to the reflection prompts per misconception

	Misconception 1	Misconception 2	Misconception 3	Misconception 4[a]
The text information contradicts the misconception	80% (24)	60% (18)	67% (20)	83% (24)
The text information completes the misconception	27% (8)	53% (16)	38% (11)	24% (7)
The text information confirms the misconception	3% (1)	30% (9)	20% (6)	7% (2)
The text information is confusing	17% (5)	7% (2)	7% (2)	21% (6)

Note. Misconception 1: the magnitude of covariance is related to the slope of the fit line in a scatter plot, Misconception 2: zero covariance shows that there is no association between two variables at all, Misconception 3: covariance proves a causal relationship between two variables, Misconception 4: covariance is a standardized statistic

[a] One participant did not complete the reflection prompts for Misconception 4. Therefore, concerning this misconception, the answers of only 29 participants could be analyzed

statement "the text information confirms the misconception" is false, because this would indicate that they recognized the conflict between the information provided in the text and the misconceptions. The results showed that these assumptions held largely true, especially for misconceptions 1 (i.e., the magnitude of covariance is related to the slope of the fit line in a scatter plot) and 4 (i.e., covariance is a standardized statistic). For misconceptions 2 (i.e., zero covariance shows that there is no association between two variables at all) and 3 (i.e., covariance proves a causal relationship between two variables), this was the case to a lesser extent. One explanation could be that the participants rather perceived that the text information complemented these misconceptions, as indicated by a relatively high agreement to the respective statement for these misconceptions. Interestingly, for misconceptions 1 and 4, the percentage of participants who stated that the text information was confusing was higher than for misconceptions 2 and 3, potentially reflecting that the participants perceived stronger cognitive conflict with regard to these misconceptions.

Furthermore, we examined whether participants' answers to the reflection prompts played a role for the effectiveness of the think sheet with reflection prompts. First, we calculated how often participants agreed to each statement across the four misconceptions (i.e., sum score from 0 to 4 for each statement). Then, we examined to what extent the outcome variables were related to the agreement scores. None of the correlations was significant ($p \geq .099$). Hence, more frequent agreement (e.g., to the statement "The text information contradicts the misconception") or less frequent agreement (e.g., to the statement "The text information confirms the misconception") to the statements of the reflection prompts did not affect the outcome variables.

Reduction of misconceptions in the groups

To examine to what extent the instructional methods reduced misconceptions, we performed a mixed ANOVA with instructional method (think sheet vs. think sheet with reflection prompts vs. control) as the between-subjects variable and time of misconceptions assessment (misconceptions test before studying vs. comprehension test) as the within-subject variable (see Fig. 5). To do so, we determined how many misconceptions the participants held not only before studying but also at the comprehension test. The ANOVA showed that there was a significant main effect of time of misconceptions assessment, $F(1, 89) = 34.79$, $p < .001$, $\eta^2 = .28$ (large effect). The main effect of instructional method failed to reach sig-

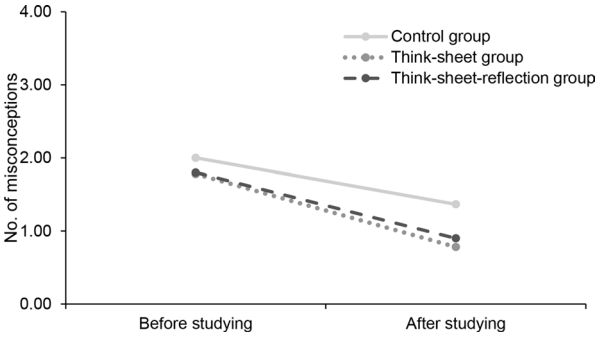


Fig. 5 Reduction of misconceptions in the three groups

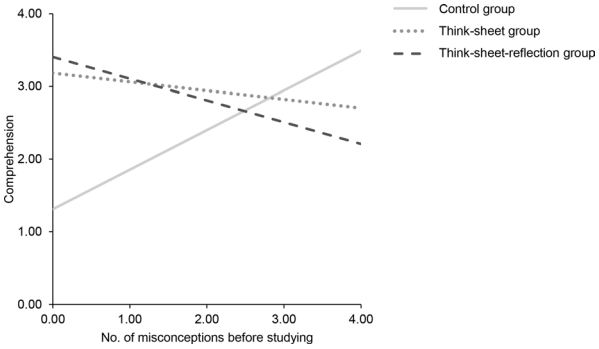


Fig. 6 Effect of misconceptions on comprehension with each method

nificance, $F(2, 89) = 2.87$, $p = .062$, $\eta^2 = .06$ (medium effect). Finally, there was no significant interaction effect between instructional method and time of misconceptions assessment, $F(2, 89) = 0.59$, $p = .559$, $\eta^2 = .01$ (small effect). This finding suggests that the number of misconceptions significantly decreased in a similar way with all three instructional methods.

Comprehension

The analysis on comprehension showed a significant overall interaction between number of misconceptions before studying and instructional method, $\Delta R^2 = .10$, $F(2, 86) = 5.03$, $p = .009$. Unexpectedly, in the control group, the number of misconceptions had a significant positive effect on comprehension, $b = 0.55$, $SE = 0.21$, $t(86) = 2.57$, $p = .012$. In contrast, the number of misconceptions had no significant effect in the think-sheet group, $b = -0.12$, $SE = 0.15$, $t(86) = -0.79$, $p = .430$, and the think-sheet-reflection group, $b = -0.30$, $SE = 0.17$, $t(86) = -1.72$, $p = .090$. Figure 6 provides a graphical representation of the conditional effects.

As can be seen in Table 5, the interactions between number of misconceptions before studying and think sheet/control and think sheet reflection/control were significant. Hence, the effect of the number of misconceptions on comprehension differed significantly between the control group and the think-sheet and think-sheet-reflection groups.

We further examined whether the impact of the number of misconceptions before studying on comprehension differed between the think-sheet group and the think-sheet-reflection group. To this end, the think-sheet group was coded as the reference group and the analysis was run again. The results showed that the interaction between number of misconcep-

Table 5 Multiple regression analysis on comprehension

Predictor variable	*b*	*SE*	*t*(86)	*p*
Constant	2.32	0.19	12.59	<.001
No. of misconceptions before studying	0.55	0.21	2.57	.012
Think sheet/control	0.64	0.26	2.49	.015
Think sheet reflection/control	0.53	0.26	2.03	.046
No. of misconceptions before studying x think sheet/control	-0.67	0.26	-2.55	.013
No. of misconceptions before studying x think sheet reflection/control	-0.84	0.28	-3.07	.003

Note. R^2=.16, $F(5, 86)$=3.16, p=.012

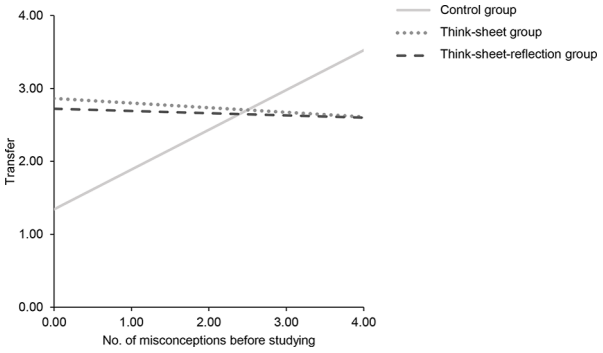


Fig. 7 Effect of misconceptions on transfer with each method

tions before studying and think sheet reflection/think sheet was not significant, b = -0.18, SE=0.23, t(86) = -0.78, p=.441.

In sum, the results were not in line with the comprehension hypothesis, because rather than students with more misconceptions before studying, those with fewer misconceptions profited from completing a think sheet. Specifically, in the control group, students who had fewer misconceptions achieved poorer comprehension compared with students who had more misconceptions. However, this effect did not occur in the think-sheet and think-sheet-reflection groups.

Transfer

In the analysis on transfer, the overall interaction between number of misconceptions before studying and instructional method failed to reach significance, ΔR^2=.06, $F(2, 86)$=2.57, p=.082. However, as in the analysis on comprehension, unexpectedly, the number of misconceptions had a significant positive effect on transfer in the control group, b=0.55, SE=0.23, $t(86)$=2.36, p=.021. In contrast, the number of misconceptions had no significant effect in the think-sheet group, b = -0.06, SE=0.17, $t(86)$ = -0.38, p=.702, and the think-sheet-reflection group, b = -0.03, SE=0.19, $t(86)$ = -0.16, p=.873. Figure 7 provides a graphical representation of the conditional effects.

As can be seen in Table 6, the interaction between number of misconceptions before studying and think sheet/control was significant, and the interaction between number of misconceptions before studying and think sheet reflection/control reached a p value of .057. Hence, there was a significant difference in the effect of the number of misconceptions on

Table 6 Multiple regression analysis on transfer

Predictor variable	*b*	*SE*	*t*(86)	*p*
Constant	2.36	0.20	11.73	<.001
No. of misconceptions before studying	0.55	0.23	2.36	.021
Think sheet/control	0.39	0.28	1.40	.166
Think sheet reflection/control	0.31	0.28	1.09	.277
No. of misconceptions before studying x think sheet/control	-0.61	0.28	-2.14	.035
No. of misconceptions before studying x think sheet reflection/control	-0.58	0.30	-1.93	.057

Note. $R^2 = .08$, $F(5, 86) = 1.43$, $p = .223$

Table 7 Multiple regression analysis on judgment accuracy of comprehension

Predictor variable	*b*	*SE*	*t*(86)	*p*
Constant	0.17	0.19	0.91	.363
No. of misconceptions before studying	-0.27	0.22	-1.26	.211
Think sheet/control	-0.38	0.26	-1.45	.150
Think sheet reflection/control	-0.27	0.26	-1.01	.314
No. of misconceptions before studying x think sheet/control	0.04	0.27	0.13	.894
No. of misconceptions before studying x think sheet reflection/control	0.35	0.28	1.24	.220

Note. $R^2 = .06$, $F(5, 86) = 1.15$, $p = .340$

comprehension between the control group and the think-sheet group, and a tendency for a difference between the control group and the think-sheet-reflection group.

We further examined whether the impact of the number of misconceptions before studying on transfer differed between the think-sheet group and the think-sheet-reflection group. The results showed that the interaction between the number of misconceptions before studying and think sheet reflection/think sheet was not significant, $b = 0.03$, $SE = 0.25$, $t(86) = 0.13$, $p = .896$.

In sum, the results were not in accordance with the transfer hypothesis, because, again, rather than students who had more misconceptions before studying, those who had fewer misconceptions profited from completing a think sheet. That is, in the control condition, students who had fewer misconceptions achieved poorer transfer performance than students who had more misconceptions. However, this effect did not occur in the think-sheet or the think-sheet-reflection group.

Judgment accuracy of comprehension

The analysis on judgment accuracy of comprehension showed that there was no significant overall interaction between number of misconceptions before studying and instructional method, $\Delta R^2 = .02$, $F(2, 86) = 1.11$, $p = .335$. Moreover, the individual interactions were not significant (see Table 7; see also Fig. 8). There was also no significant interaction between number of misconceptions before studying and think sheet reflection/think sheet, $b = 0.31$, $SE = 0.24$, $t(86) = 1.32$, $p = .190$. Running the regression without the interaction terms yielded no significant main effects. Overall, the results appeared to be not in line with the comprehension-judgment hypothesis.

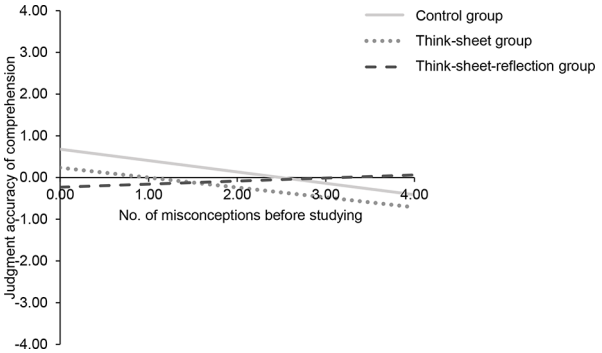


Fig. 8 Effect of misconceptions on judgment accuracy of comprehension with each method

Table 8 Multiple regression analysis on judgment accuracy of transfer

Predictor variable	*b*	*SE*	*t*(86)	*p*
Constant	-0.37	0.19	-1.90	.061
No. of misconceptions before studying	-0.46	0.22	-2.03	.045
Think sheet/control	-0.18	0.27	-0.68	.498
Think sheet reflection/control	-0.07	0.27	-0.25	.802
No. of misconceptions before studying x think sheet/control	0.19	0.28	0.67	.503
No. of misconceptions before studying x think sheet reflection/control	0.38	0.29	1.30	.198

Note. $R^2 = .08$, $F(5, 86) = 1.47$, $p = .207$

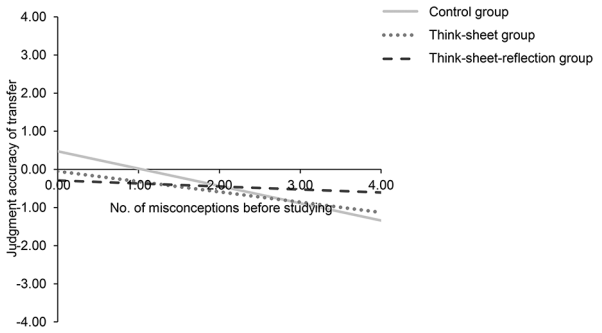


Fig. 9 Effect of misconceptions on judgment accuracy of transfer with each method

Judgment accuracy of transfer

The analysis on judgment accuracy of transfer also revealed no significant overall interaction between number of misconceptions before studying and instructional method, $\Delta R^2 = .02$, $F(2, 86) = 0.86$, $p = .427$. Moreover, the individual interactions were not significant (see Table 8; see also Fig. 9). There was also no significant interaction between number of misconceptions before studying and think sheet reflection/think sheet, $b = 0.19$, $SE = 0.24$, $t(86) = 0.78$, $p = .435$. Running the regression without the interaction terms yielded a significant negative main effect of number of misconceptions, $b = -0.25$, $SE = 0.11$, $t(88) = -2.34$, $p = .022$. This means that, in general, learners who had more misconceptions before studying more strongly underestimated their transfer performance. Overall, the results were not in accordance with the transfer-judgment hypothesis.

Table 9 Multiple regression analysis on regulation decision adequacy

Predictor variable	*b*	*SE*	*t*(86)	*p*
Constant	0.60	0.05	13.07	<.001
No. of misconceptions before studying	0.00	0.05	0.00	1.00
Think sheet/control	0.05	0.06	0.78	.435
Think sheet reflection/control	0.02	0.07	0.34	.736
No. of misconceptions before studying x think sheet/control	0.02	0.07	0.28	.780
No. of misconceptions before studying x think sheet reflection/control	-0.05	0.07	-0.78	.437

Note. $R^2 = .03$, $F(5, 86) = 0.47$, $p = .799$

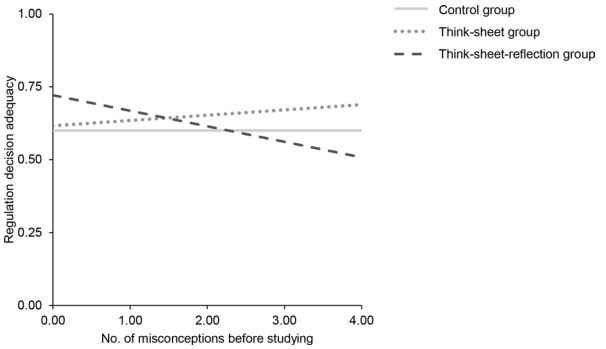


Fig. 10 Effect of misconceptions on regulation decision adequacy with each method

Regulation decision adequacy

The analysis on regulation decision adequacy showed no significant overall interaction between number of misconceptions before studying and instructional method, $\Delta R^2 = .02$, $F(2, 86) = 0.80$, $p = .453$. Moreover, the individual interactions were not significant (see Table 9; see also Fig. 10). There was also no significant interaction between number of misconceptions before studying and think sheet reflection/think sheet, $b = -0.07$ $SE = 0.06$, $t(86) = -1.25$, $p = .216$. Running the regression without the interaction terms yielded no significant main effects. To conclude, the results did not support the regulation-adequacy hypothesis.

Discussion

The goal of the present study was to investigate the effectiveness of think sheets with and without the addition of reflection prompts as a method to revise statistical misconceptions and support self-regulated learning from text. Overall, the results showed that for students with more misconceptions before studying, completing a think sheet with or without reflection prompts provided no significant benefit compared to solely reading about the misconceptions and correct information. Furthermore, no differences emerged between completing a think sheet with and without reflection prompts. Unexpectedly, regarding comprehension and transfer, the results showed that in the reading-only condition, students with fewer misconceptions before studying performed worse than those with more misconceptions – an effect that did not occur with think sheets. Judgment accuracy of comprehension and regulation decision adequacy were unaffected by the number of misconceptions and/or the

instructional method used. Regarding judgment accuracy of transfer, students with more misconceptions before studying more strongly underestimated their transfer ability, regardless of the instructional method.

Reduction of misconceptions

As was expected based on prior research (e.g., Prinz et al., 2018, 2019, 2021; Huck, 2016), there was quite a high prevalence of misconceptions about the statistical concept of covariance among the students. This reinforces that instructional methods are needed to address these misconceptions. Accordingly, a notable finding for instruction is that the number of misconceptions students held decreased when they completed a think sheet without reflection prompts, completed a think sheet with reflection prompts, or only read about the misconceptions and correct information but did not complete a think sheet. It can therefore be assumed that all three methods effectively elicited some conceptual-change processes. This result coincides with the finding of Prinz et al. (2021) that completing a think sheet, reading a refutation text, and just reading about the misconceptions and correct information were comparably effective in reducing misconceptions. In the present study and the study by Prinz et al. (2021), the applied instructional methods shared a common element: They presented students with both the misconceptions and the correct information (instead of only the correct information). Apparently and in line with the KReC framework (Kendeou & O'Brien, 2014), when learners are explicitly confronted with both the misconceptions and the respective correct information, the two become coactivated in learners' memories, enabling the experience of cognitive conflict and further conceptual-change processes (cf. Kendeou & O'Brien, 2014).

Another not mutually exclusive explanation for why misconceptions were reduced with all three methods could be that the misconceptions were not strongly endorsed by the students. Although participants were required to have attended at least one statistics course that covered covariance, they might not have possessed a very deep understanding of the topic. Thus, the students might not have heavily challenged the textual information, making conceptual change relatively easy to achieve. For more strongly held misconceptions, a different picture might emerge. Specifically, in this case, there might be more resistance to conceptual change, and completing a think sheet might turn out as a more effective method. This interpretation is in line with the findings by van Loon et al. (2015). They showed that students corrected misconceptions held with low confidence to a similar degree by reading refutation texts and standard texts. In contrast, students more frequently corrected misconceptions held with high confidence after reading refutation texts than after reading standard texts.

Comprehension and transfer

Concerning comprehension and transfer, the results did not show the expected interaction effect that students with more misconceptions before studying would perform worse than students with fewer misconceptions when solely reading about the misconceptions and correct information but not when completing a think sheet with or without reflection prompts. Rather, the results indicated that students with more misconceptions did not profit more from completing a think sheet than from reading alone. This finding contradicts the study

by Prinz et al. (2021), which showed that compared to reading only, completing a think sheet benefited the comprehension and transfer performance of students with more misconceptions. It should be noted, though, that in the present study, the reading-only condition yielded quite strong effects for students who had a higher number of misconceptions before studying. That is, unexpectedly, students with more misconceptions who only read about the misconceptions and correct information achieved quite high performance levels, which may have constrained the potential to observe distinct effects resulting from the instructional methods. Nonetheless, there was room for improvement in the think-sheet conditions, and the comprehension and transfer performance of students with more misconceptions could have been more strongly supported, as would be expected based on theory. According to the KReC framework (Kendeou & O'Brien, 2014), effective knowledge revision requires the *coactivation* of misconceptions and correct information in learners' memories, after which the *integration* of the new correct information into existing knowledge structures can take place. In this study, think sheets may not have sufficiently supported these processes of knowledge revision, particularly integration. As noted previously, all three methods likely promoted the coactivation of the misconceptions and correct information. However, completing a think sheet only to a limited extent may have fostered deeper cognitive engagement necessary for the effective integration of the new information. Thus, completing a think sheet may not have yielded the anticipated cognitive benefits in this study. This outcome calls into question the assumed effectiveness of think sheets and highlights the need for further exploration of the mechanisms influencing their impact. For example, learners' cognitive load might play a role for the effectiveness of think sheets. Potentially, when think sheets induce high cognitive load because they are complex and require too much information processing at once, this could hinder learners' ability to integrate the new information effectively. Investigating such boundary conditions could help expand and refine existing theories such as the KReC framework (Kendeou & O'Brien, 2014). For example, it would allow researchers to pose and systematically test theoretical assumptions about how (e.g., think sheets, reading only) and under what conditions (e.g., level of cognitive load) processes such as coactivation and integration can be most effectively achieved.

As indicated above and contrary to expectations, the results showed that in the reading-only condition, students who had more misconceptions before studying achieved better comprehension and transfer performance than those who had fewer misconceptions. This means that students whose knowledge was contaminated with misconceptions profited from reading about the misconceptions and correct information, whereas students who had largely correct prior knowledge were disadvantaged by this method. Research has shown that people are more likely to believe in information when they perceive that many other people believe in it, too (e.g., Lewandowsky et al., 2012). Accordingly, the way the misconceptions were introduced in the instructional materials (i.e., concerning the think and control sheet), namely, as typical false beliefs that learners have about these aspects of covariance, may have contributed to this effect. Specifically, when students who had largely correct knowledge in the beginning were exposed to the misconceptions and correct information but without any further task, they might have concluded that, because the misconceptions are widespread, there must be something to them or at least become confused. Consequently, their comprehension and transfer performance were impaired. In contrast, when students completed a think sheet, this effect did not occur, and comprehension and transfer performance were not affected by misconceptions. Presumably, when students who

had largely correct knowledge in the beginning were not only exposed to the misconceptions and correct information but also had to engage in the generative task of completing a think sheet, they were better able to recognize the incorrectness of the misconceptions and acquire the correct information. In turn, their comprehension and transfer performance were not impaired. To conclude, the present study suggests that when learners have some misconceptions, all three methods are adequate to facilitate their comprehension and knowledge transfer. However, when learners have no or almost no misconceptions, completing a think sheet seems more beneficial than solely reading about the misconceptions and correct information.

One might argue that the results could also be explained by students with fewer misconceptions in the control group having more other incorrect prior knowledge (i.e., other than the misconceptions), which in turn led to poorer learning outcomes. However, this potential explanation is not supported, as there was no significant association between the number of misconceptions and the number of incorrect responses in the misconceptions test for this group (Spearman's $r = -.20$, $p = .297$). Hence, students with fewer misconceptions did not have more other incorrect prior knowledge. In addition, it should be noted that other incorrect prior knowledge than the misconceptions was not very common among the students (22 participants gave no incorrect responses, seven participants gave one incorrect response, and one participant gave two incorrect responses). Instead, they mainly held the misconceptions or the correct knowledge.

It is striking that prior research found the reverse effect of misconceptions when students engaged in reading only. That is, students who had more misconceptions before studying, rather than fewer misconceptions, tended to achieve poorer comprehension and transfer performance when reading about the correct information (Prinz et al., 2018, 2019) or when reading about the correct information and the misconceptions (Prinz et al., 2021). In the present study, we implemented a strong control group that did not differ with regard to the amount of information received but only with regard to the learning task. Specifically, we provided participants in the control group not only with the correct information but also with the misconceptions. It remains unclear how the students would have performed if they had read only about the correct information and not about the misconceptions. Potentially, in this case, no performance detriment for students with fewer misconceptions would have occurred (cf. Prinz et al., 2018, 2019). Further research is needed that clarifies the conditions under which misconceptions have a negative or positive influence on learning when students are provided with both the correct information and the misconceptions to explain the contradictory results of the study by Prinz et al. (2021) and the present study. Nonetheless, it is important to emphasize that think sheets proved beneficial on both occasions: They compensated either for the negative (Prinz et al., 2021) or the positive (present study) influence of misconceptions on learning. In other words, compared to reading only, completing a think sheet was beneficial either for those who had more (Prinz et al., 2021) or those who had fewer misconceptions (present study) before studying. However, given the conflicting findings between the two studies, no definitive conclusions can be drawn about the role of think sheets in addressing statistical misconceptions. As indicated previously, more research is needed to explore the conditions under which think sheets are most beneficial. This implies that no clear guidelines for educators regarding the implementation of think sheets can yet be derived. At this stage of research, it seems most advisable for educators to consider think sheets as one potential tool when working on topics where misconceptions

exist, adapting their use based on students' needs and further emerging evidence. Educators should also monitor and evaluate the impact of think sheets in their classrooms, adjusting their use depending on student outcomes.

Although, prior research showed that reflection prompts can promote knowledge acquisition and transfer (e.g., Bannert, 2006; Lin & Lehman, 1999), the think sheet with reflection prompts was not superior to the think sheet without reflection prompts with regard to comprehension and transfer in the present study. There are several potential explanations for why the reflection prompts did not produce additional benefits. One explanation is that answering the prompts did not induce further cognitive processes than those already elicited by completing the think sheet. More precisely, by requiring students to extract the correct information from the text and to set it against the respective misconceptions, completing the think sheet likely evoked coactivation and some further conceptual-change processes (e.g., Prinz et al., 2021). Answering the reflection prompts required students to contrast the misconceptions and the correct information, which might again have stimulated coactivation and some further conceptual-change processes. However, because these processes had already been elicited when answering the think sheet, the effects of answering the reflection prompts might have been too small to lead to a further advantage.

Another explanation is related to how the prompts were implemented. First, when completing the prompts for each block, the students ideally should have indicated that the statement "the text information contradicts the misconception" is true and, conversely, that the statement "the text information confirms the misconception" is false. However, the students might have perceived always giving the same answer to the prompts as odd. As a result, the prompts might have been confusing to some extent, preventing an added value. Second, the prompts were specific statements that provided clear instructions about what to think. Previous research showed that generic prompts asking students to "stop and think" elicited more productive reflection and led to a more coherent understanding than directed prompts that gave students clear hints about what to think (Davis, 2003). Therefore, generic prompts that, for example, ask students about their thoughts concerning the relation of the misconceptions and correct information might yield more favorable effects. At the same time, such prompts would prevent the problem that statements that always require the same answer can confuse participants.

A final explanation is that the prompts exerted effects during learning, but that these effects did not necessarily translate into superior comprehension and transfer performance. This assumption is supported by the study by Theobald et al. (2024). They found that children who received reflection prompts when making predictions were better able to recognize incorrect predictions, more likely to subsequently make correct predictions, and faster to improve their performance during learning. However, no effects of the prompts on posttest and transfer test performance after learning occurred. Similarly, in the present study, it might have been the case that the reflection prompts supported the students in monitoring and recognizing conflict as well as in integrating correct information more quickly while studying. Yet, effects while studying were not assessed. Finally, it needs to be emphasized that although think sheets with and without reflection prompts showed similar effects, the results support the KReC framework (Kendeou & O'Brien, 2014), which emphasizes the importance of coactivation and further conceptual-change processes for enhancing learning (see, e.g., also van den Broek & Kendeou, 2008).

Judgment accuracy and regulation decision adequacy

The accuracy of students' judgments of comprehension did not depend on how many misconceptions they had and/or on the instructional method they received. In previous research, under standard-reading conditions, students with more misconceptions before studying overestimated their comprehension more strongly than students with fewer misconceptions (e.g., Prinz et al., 2018, 2019, 2021). In the present study, this detrimental influence of misconceptions when reading only did not emerge, potentially constraining the observation of a compensatory effect of completing a think sheet. It appeared that, in general, the students were quite accurate in judging their comprehension. Hence, it seems that all instructional methods were beneficial to support students' judgment accuracy of their comprehension. Prinz et al. (2021) suggested that coactivation represents an important precondition not only for achieving correct comprehension but also for achieving high judgment accuracy. Specifically, in the context of coactivation and further conceptual-change processes, learners might inevitably monitor and evaluate their understanding more carefully (Prinz et al., 2021). As previously indicated, all three methods in the present study might have induced coactivation and some further conceptual-change processes, as misconceptions were effectively reduced with all of them. In turn, the students' monitoring and evaluation of their comprehension might have been supported with all three methods to a similar extent.

Concerning judgment accuracy of transfer, Prinz et al. (2021) found that when completing a think sheet, misconceptions tended to have a negative effect. That is, students who had more misconceptions before studying more strongly underestimated their ability to transfer their acquired comprehension. In the present study, a negative effect of misconceptions occurred not only when completing a think sheet, but in general. Specifically, there was a significant negative main effect of misconceptions, indicating that students who had a higher number of misconceptions before studying more strongly underestimated their transfer performance. Given these results, it appears that, concerning transfer in particular, underconfidence is an issue. A potential explanation is that students consider transferring their comprehension to be more difficult than achieving it, but overrate the difference in difficulty between the two tasks (cf. Prinz et al., 2021). For students who have more misconceptions before studying, this tendency to underestimate their transfer performance might be especially pronounced.

With regard to regulation decision adequacy, no effects of misconceptions and/or the instructional method emerged. Previous research showed that more accurate judgments of one's comprehension can lead to more adequate regulation decisions (e.g., Rawson et al., 2011; Thiede et al., 2003). In the present study, all three instructional methods supported students in accurately judging their comprehension. In turn, their regulation decisions might have been promoted independently of the received method as well. Overall, students' regulation decisions were of medium adequacy with room for further improvement. Therefore, additional means to help students make adequate decisions about their further learning would be beneficial. For example, prompts that ask students to think about why or why not they should engage in restudy activities could be implemented within think sheets.

Limitations and future directions

First, as indicated previously, we did not assess effects of completing a think sheet and answering reflection prompts on processes during learning. For example, engaging with reflection prompts might facilitate conflict monitoring and faster knowledge integration while studying. Therefore, further research should shed light on these "online" learning processes, for example, by means of response times or pupil dilation responses (cf. Theobald et al., 2024).

Second, think sheets and prompts can be implemented in various ways. For instance, rather than providing misconceptions on a think sheet, which is a more passive approach to activate potential misconceptions, learners can be asked to enter their own preexistent beliefs (cf. Dole & Smith, 1989). This approach requires learners to deliberately activate their prior beliefs before being confronted with the correct information, which may be even more effective for inducing coactivation and conceptual-change processes through think sheets. In addition, as stated before, compared to specific reflection prompts that guide learners towards very specific cognitive activities, generic reflection prompts that allow for varying thought processes might yield greater effects. Thus, future studies should compare the effectiveness of specific and generic reflection prompts within think sheets. Examining the role of design features such as the activation of prior beliefs and the specificity of reflection prompts can provide deeper insights into the conditions under which think sheets and reflection prompts are more or less effective. This, in turn, would help refine and expand the theoretical bases that underpin their use (e.g., Bannert, 2009; Kendeou & O'Brien, 2014).

Third, we examined how thoroughly the participants completed the think sheets and reflection prompts. The results suggested that the participants approached the tasks seriously and completed them with high engagement. In addition, the results showed that the quality with which the think sheets were completed (i.e., more correct and complete answers) played a role for their effectiveness. Hence, helping students to complete think sheets with the correct and complete information could increase positive effects of this instructional method. Thus, further research might add feedback to think sheets and examine its impact.

Fourth, it cannot be excluded that social desirability played a role for participants' responses, potentially biasing the results. For example, participants might have indicated that they would like to restudy just because they felt obligated to do so when taking part in a study. To mitigate this possibility, the instructions explicitly asked participants whether they would like to restudy in order to improve their comprehension and better prepare for the upcoming test questions. Even so, participants might have made their decisions due to other reasons. Similar limitations might apply to the judgments. For instance, participants might have provided lower judgments than they would have in real-life situations to avoid appearing presumptuous. Conducting field studies in more natural classroom environments may help to reduce these issues. Moreover, when responding to the reflection prompts, participants might have based their answers on what they thought the task would require. For example, they might have indicated that the statement "The text information contradicts the misconception" is true simply because they presumed that the text would provide the correct information, without thoroughly grasping how it contradicted the misconception. However, it should be noted that the agreement rates regarding the statements were not absolute (i.e., not 100% or 0%; see Table 4). For instance, the agreement with the statement "The text information contradicts the misconception" ranged from 60% to 83% for the four miscon-

ceptions. This suggests that participants did not merely choose responses they believed were expected (e.g., "true"), but at least to some extent considered how the textual information aligned with the misconceptions. Nonetheless, future studies could include misconceptions that are not refuted by the text or include correct beliefs that are confirmed by the text to assess how participants respond in these cases and identify potential response biases.

Fifth, there were significant group differences concerning the number of attended statistics courses that covered covariance. However, the majority of participants in all three groups had attended between one and two such courses, with only a few participants having attended more or fewer. Moreover, aside from the number of statistics courses on covariance, no other a priori differences were observed concerning relevant variables, such as the number of misconceptions about covariance or the number of semesters. While we cannot entirely rule out the possibility of initial group differences in unmeasured variables, such as attitude towards statistics, the random assignment of participants to groups makes it unlikely that such differences existed and systematically influenced the results.

Finally, it would be interesting to examine to what extent the results of the present study can be transferred to other domains, misconceptions, and age groups. For example, in this study, we did not assess the confidence with which the misconceptions were endorsed by the students, and the results may vary for misconceptions held with different degrees of confidence. Presumably, compared with low-confidence misconceptions, high-confidence misconceptions might be more resistant to change and more strongly impair learning. In this case, the instructional method of completing a think sheet might prove more beneficial (cf. van Loon et al., 2015). Moreover, school students usually need more instructional guidance compared with university students who are typically better able to judge and regulate their own learning (Prinz et al., 2020). For this reason, younger school students might profit from think sheets and reflection prompts more strongly than older university students. Hence, it would be insightful to explore whether different age groups benefit from think sheets and reflection prompts differently.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data are available from the corresponding author on request.

Declarations

Ethics guidelines The study was conducted in alignment with the ethical principles of the American Psychological Association and approved by the institutional ethics committee.

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

References

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research*, *35*(4), 359–375. https://doi.org/10.2190/94V6-R58H-3367-G388

Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift Für Pädagogische Psychologie*, *23*(2), 139–145. https://doi.org/10.1024/1010-0652.23.2.139

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, *40*(1), 193–211. https://doi.org/10.1007/s11251-011-9167-4

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, *52*(1), 293–306. https://doi.org/10.1016/j.chb.2015.05.038

Bascandziev, I. (2020). Inconsistencies among beliefs as a basis for learning via thought experiments. In B. C. Armstrong, S. Denison, M. Mack, & Y. Xu (Hrsg.), *Proceedings of the 42nd annual conference of the cognitive science society* (S. 437–442). Cognitive Science Society.

Bascandziev, I. (2024). Thought experiments as an error detection and correction tool. *Cognitive Science*, *48*(1). https://doi.org/10.1111/cogs.13401. Article e13401.

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und sozialwissenschaftler [Statistics for human and social scientists]* (7th ed.). Springer.

Braumann, S., van Wermeskerken, M. M., van de Pol, J., Pijeira-Díaz, H. J., de Bruin, A. B., & van Gog, T. (2024). The role of feedback on students' diagramming: Effects on monitoring accuracy and text comprehension. *Contemporary Educational Psychology*, *76*, 102251. https://doi.org/10.1016/j.cedpsych.2023.102251

Brummer, T., & Macceca, S. (2013). *Reading strategies for mathematics*. Shell Education.

Danielson, R. W., Jacobson, N. G., Patall, E. A., Sinatra, G. M., Adesope, O. O., Kennedy, A. A. U., & Sunday, O. J. (2024). The effectiveness of refutation text in confronting scientific misconceptions: A meta-analysis. *Educational Psychologist*. Advance online publication. https://doi.org/10.1080/00461520.2024.2365628

Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences*, *12*(1), 91–142. https://doi.org/10.1207/S15327809JLS1201_4

Davis, E. A., & Linn, M. C. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education*, *22*(8), 819–837. https://doi.org/10.1080/095006900412293

Dole, J. A., & Smith, E. L. (1989). Prior knowledge and learning from science text: An instructional study. In S. McCormick & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction. Thirty-eighth yearbook of the National Reading Conference* (S. 345–352). National Reading Conference.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, *52*(4), 551–565. https://doi.org/10.1016/j.jml.2005.01.011

Fang, Z., Lamme, L. L., & Pringle, R. M. (2010). *Language and literacy in inquiry-based science classrooms, grades 3–8*. Corwin.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221–233. https://doi.org/10.1037/h0057532

Golke, S., Steininger, T., & Wittwer, J. (2022). What makes learners overestimate their text comprehension? The impact of learner characteristics on judgment bias. *Educational Psychology Review*, *43*(4), 2405–2450. https://doi.org/10.1007/s10648-022-09687-0

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.

Huck, S. W. (2016). *Statistical misconceptions*. Routledge.

Johnson, A. M., Azevedo, R., & D'Mello, S. K. (2011). The Temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, *29*(4), 471–504. https://doi.org/10.1080/07370008.2011.610244

Kendeou, P., & O'Brien, E. J. (2014). The knowledge revision components (KReC) framework: Processes and mechanisms. In D. N. Rapp, & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational science* (pp. 353–377). MIT University.

Kendeou, P., & Van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, *97*(2), 235–245. https://doi.org/10.1037/0022-0663.97.2.235

Kramarski, B., & Michalsky, T. (2013). Student and teacher perspectives on IMPROVE self-regulation prompts in web-based learning. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (S. 35–51). Springer. https://doi.org/10.1007/978-1-4419-5546-3_3

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, *36*(7), 837–858. https://doi.org/10.1002/(SICI)1098-2736(199909)36:7%3C837::AID-TEA6%3E3.0.CO;2-U

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

McMaster, K. L., & Kendeou, P. (2023). Refocusing reading comprehension: Aligning theory with assessment and intervention. *Learning and Individual Differences, 102*, Article 102256. Advance online publication. https://doi.org/10.1016/j.lindif.2023.102256

McNamara, D. S. (2007). *Reading comprehension strategies: Theories, interventions, and technologies*. Psychology Press.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26). Academic.

Organisation for Economic Co-operation and Development. (2006). *PISA 2006–Schulleistungen im internationalen Vergleich: Naturwissenschaftliche Kompetenzen für die Welt von morgen* [PISA 2006: Science competencies for tomorrow's world] (Vol. 1). Retrieved from https://www.oecd.org/content/dam/oecd/de/publications/reports/2007/12/pisa-2006_g1gh866e/9789264041257-de.pdf

Prinz, A., Golke, S., & Wittwer, J. (2018). The double curse of misconceptions: Misconceptions impair not only text comprehension but also metacomprehension in the domain of statistics. *Instructional Science, 46*(5), 723–765. https://doi.org/10.1007/s11251-018-9452-6

Prinz, A., Golke, S., & Wittwer, J. (2019). Refutation texts compensate for detrimental effects of misconceptions on comprehension and metacomprehension accuracy and support transfer. *Journal of Educational Psychology, 111*(6), 957–981. https://doi.org/10.1037/edu0000329

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, 31*, Article 100358. Advance online publication. https://doi.org/10.1016/j.edurev.2020.100358

Prinz, A., Golke, S., & Wittwer, J. (2021). Counteracting detrimental effects of misconceptions on learning and metacomprehension accuracy: The utility of refutation texts and think sheets. *Instructional Science, 49*(2), 165–195. https://doi.org/10.1007/s11251-021-09535-8

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

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

Smith, P. L., & Ragan, T. J. (2005). *Instructional design* (3rd ed.). Wiley.

Stark, R., & Krause, U. M. (2009). Effects of reflection prompts on learning outcomes and learning behaviour in statistics education. *Learning Environments Research*, *12*(3), 209–223. https://doi.org/10.1007/s10984-009-9063-x

Theobald, M., Colantonio, J., Bascandziev, I., Bonawitz, E., & Brod, G. (2024). Do reflection prompts promote children's conflict monitoring and revision of misconceptions? *Child Development*. https://doi.org/10.1111/cdev.14081. Advance online publication.

Thiede, K. W., Anderson, M. C. M., & Therriault, 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., 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

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., van Gog, T., Braumann, S., & de Bruin, A. (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*, *32*(4), 951–977. https://doi.org/10.1007/s10648-020-09560-y

Van de Pol, J., van Gog, T., & Thiede, K. (2021). The relationship between teachers' cue-utilization and their monitoring accuracy of students' text comprehension. *Teaching and Teacher Education*, *107*, 103482. https://doi.org/10.1016/j.tate.2021.103482

Van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, *22*(3), 335–351. https://doi.org/10.1002/acp.1418

Van Loon, M. H., Dunlosky, J., van Gog, T., van Merriënboer, J. J. G., & de Bruin, A. B. H. (2015). Refutations in science texts lead to hypercorrection of misconceptions held with high confidence. *Contemporary Educational Psychology*, *42*, 39–48. https://doi.org/10.1016/j.cedpsych.2015.04.003

Wirth, J. (2009). Promoting self-regulated learning through prompts. *Zeitschrift Für Pädagogische Psychologie*, *23*(2), 91–94. https://doi.org/10.1024/1010-0652.23.2.91

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerls, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic.