RESEARCH ARTICLE

WILEY

Caption it! The impact of headings on learning from texts

Anja Prinz-Weiß[1] | Aline König[2]

[1]Department of Psychology, University of Education Karlsruhe, Karlsruhe, Germany

[2]University of Education Freiburg, Freiburg im Breisgau, Germany

Correspondence
Anja Prinz-Weiß, Department of Psychology, University of Education Karlsruhe, Erzbergerstr. 119, 76133 Karlsruhe, Germany.
Email: anja.prinz-weiss@ph-karlsruhe.de

Abstract

When learning from texts, it is not only important that learners remember and comprehend the content, but also that they monitor and accurately judge their memory and comprehension so as to efficiently regulate their learning. In the present experiment with 51 university students, we investigated to what extent headings within texts promote these processes. The results revealed that headings supported learners in comprehending the texts as well as in accurately judging their comprehension. The effects of headings on memory and judgment accuracy concerning memory were not significant. Moreover, headings affected learners' cognitive load. This study indicates the usefulness of including headings in texts to support learners' self-regulated learning from texts.

KEYWORDS
cognitive load, headings, judgment accuracy, learning from texts

1 | INTRODUCTION

Learning by reading texts is one of the most common means through which knowledge is acquired. When learning from texts, it is important that learners remember and comprehend the content. In addition, they must monitor and accurately judge their memory and comprehension (e.g., McNamara & Magliano, 2009; Thiede et al., 2003). Including headings within texts might help to promote these cognitive and metacognitive processes, because they signal the structure of a text. In the present study, we investigated to what extent headings within texts improve learners' memory and comprehension as well as their judgment accuracy concerning their memory and comprehension. Moreover, we explored effects of headings on learners' cognitive load.

1.1 | The impact of headings on memory and comprehension

When learners study texts, they typically not only need to remember important details but also to draw inferences and gain deeper comprehension. For example, a learner who reads a text on the circulatory system should be able to remember crucial terms and their meanings, such as what veins and arteries are, as well as be able to explain the dynamic process of blood flow throughout the human body. However, good memory and deep comprehension of a text are not always easily achieved (e.g., Perfetti et al., 2005).

Research has shown that headings within texts can facilitate memory for text content. Specifically, headings resulted in a greater number of textual idea units and topics mentioned in free recall as well as in enhanced remembrance of facts reported in a text (e.g., Hartley & Trueman, 1985; Holley et al., 1981; Lorch Jr. & Lorch, 1996; Sanchez et al., 2001; Wilhite, 1986). In addition, research has indicated that headings can support the comprehension of text content (e.g., Hyönä & Lorch, 2004; Mautone & Mayer, 2001; Spyridakis & Standal, 1987; Surber & Schroeder, 2007; Wilhite, 1986). For example, headings led to an increased number of core points regarding different topics mentioned in learners' summaries of a text's main content (Hyönä & Lorch, 2004). In accordance with these earlier findings, more recent research on the signaling principle also indicates that learning is improved when relevant elements or the organization of the instructional material is emphasized. For example, apart from headings, visual

signals within texts such as highlighting were found to be beneficial (e.g., Richter et al., 2016; Schneider et al., 2018).

Theoretically, it has been assumed that during online processing, headings trigger learners' prior knowledge, which guides their comprehension. In addition, headings help learners create a mental model of the text's topic structure for organizing and integrating the information. During output processing, such a topic-structure model renders the text topics and subordinate information easier to retrieve from memory (e.g., Holley et al., 1981; Hyönä & Lorch, 2004; Lemarié et al., 2012; Lorch et al., 2001). Support for the assumed processes elicited by headings during online processing has been provided by eye-tracking studies. For example, Hyönä and Lorch (2004) found that readers devoted particular attention to headings. In addition, headings facilitated the processing of sentences introducing a new topic and reduced the need to look back at such topic sentences. Furthermore, Cauchard et al. (2010) revealed that headings supported search processes for relevant information within a text. These findings indicate that headings are salient signals that activate prior knowledge and facilitate further processing.

The assumed effects of headings should also be reflected in learners' cognitive load. According to the cognitive load theory (e.g., Sweller et al., 1998), working memory capacity is limited and divided between different sources of load. Traditionally, three types of load have been distinguished: Intrinsic load refers to the subjective complexity of the to-be-learned material. Extraneous load is unproductive load that results from poor instructional design. Germane load arises from processes directly contributing to learning.[1] It can be assumed that headings enhance learning-relevant germane load by supporting the activation of relevant prior knowledge as well as organization and integration processes, while reducing extraneous load by making the topic structure more salient. Britton et al. (1982) assessed participants' reaction times to a secondary task performed while reading signaled or non-signaled texts. Signaled texts contained previews in the form of titles or headings, relational words (e.g., "consequently"), importance phrases (e.g., "an essential point is"), and summaries. The results revealed that secondary-task reaction times were faster with signaled than with non-signaled texts, suggesting that the processing of texts with signals required less cognitive capacity. Beege et al. (2020) examined how highlighting the main concepts in a text in red affects learners' intrinsic and extraneous load. The results showed that the visual signals had no effect on intrinsic load but led to reduced extraneous load, indicating that the learners were distracted by irrelevant information to a lesser extent. Finally, a meta-analysis on the signaling principle revealed that signaled material generally resulted in diminished cognitive load. As this effect was rather small, it was interpreted as indicating that while signaling reduces extraneous load, it increases germane load to some extent (Schneider et al., 2018). However, these studies did not or not exclusively focus on headings, but manipulated other or several signaling features, so that the specific effect of headings on cognitive resources during reading remains unclear. In addition, none of the studies examined intrinsic, extraneous, and germane load, limiting specific conclusions concerning the three types of load. Therefore, to improve our understanding of how headings contribute to learning, it is important to further investigate their impact on learners' cognitive load.

1.2 | The impact of headings on metamemory and metacomprehension accuracy

It is crucial that learners accurately monitor and judge their memory and comprehension of texts. The former is referred to as metamemory accuracy and the latter as metacomprehension accuracy (e.g., Griffin et al., 2019; Wiley et al., 2008). Accurate judgments are important because they support the effective regulation of one's studying and thus successful learning (e.g., Rawson et al., 2011; Thiede et al., 2003). For example, learners who accurately judge that they have not yet learned the text content well can selectively devote further time and efforts to this content. Two of the most common measures to assess the accuracy of learners' judgments are bias and relative accuracy. Bias is calculated by subtracting an individual's actual from judged performance, thus indicating over- or underconfidence. Relative accuracy is operationalized as the intraindividual correlation between an individual's judgments and actual performance scores, hence reflecting the extent to which a learner accurately differentiates between better and worse learned material (e.g., Schraw, 2009). Research has shown that learners often exhibit poor accuracy in terms of both bias and relative accuracy. Specifically, they tend to overestimate their learning and are poor at discriminating between better and worse learned material (e.g., Maki et al., 2005; Wiley et al., 2016; see also Prinz et al., 2020a). The cue utilization framework (Griffin et al., 2009; cf. Koriat, 1997) suggests that a reason for this poor accuracy is that learners often use inappropriate cues when making judgments. For example, they often base their judgments on heuristic cues, such as their familiarity with the domain, which usually do not closely relate to their actual performance with respect to a specific text.

Whether headings within texts have the potential to enhance metamemory and metacomprehension accuracy is unclear so far. However, previous research has indicated that features of the texts learners read can play a role for their judgment accuracy. For instance, metacomprehension accuracy tends to be higher for texts on multiple, diverse topics (e.g., black holes, blood sugar, evolution, sea levels, and viruses) than for texts on a single topic (e.g., Prinz et al., 2020a). This finding suggests that reading texts on different topics makes potential contrasts between them more obvious, increasing learners' awareness that their comprehension of the texts might differ. Consequently, they might more carefully monitor their learning and use more valid cues when making judgments (cf. Prinz et al., 2020a). Similarly, headings indicating the different topics covered within a text might increase learners' monitoring efforts and judgment accuracy. Moreover, it has been found that learners' judgment accuracy can be enhanced when relevant information within texts is highlighted in color (e.g., Gier et al., 2010). This highlighting might direct learners' monitoring to the relevant parts of a text. Likewise, headings might emphasize important content, which promotes learners' monitoring of it and hence their judgment accuracy.

Furthermore, it has been suggested that cognitive load also plays a role in monitoring activities (e.g., de Bruin et al., 2020; Seufert, 2018). The demands put on learners' cognitive system leave a certain amount of resources available for monitoring. Thus, by reducing extraneous load, headings might enhance not only learning-relevant cognitive processes but also monitoring processes. Specifically, by making it easier for learners to mentally represent the topic structure of a text, headings might free cognitive capacity for monitoring.

1.3 | The present study

In the present study, we examined the effects of headings within texts on learners' memory and comprehension as well as on their judgment accuracy concerning their memory and comprehension. The study extends prior research in several important ways. First, we assessed the impact of headings on learners' ability to remember details as well as to draw inferences on the same texts. Second, alongside the effects of headings on cognitive processes, we examined their effects on metacognitive processes by using two measures of monitoring accuracy, namely bias and relative accuracy. Third, we explored the mechanisms underlying the effects of headings by considering learners' cognitive load.

The following hypotheses were investigated: In line with previous studies on the effects of headings on learning (e.g., Hartley & Trueman, 1985; Hyönä & Lorch, 2004), we expected to find positive effects of headings on learners' memory (memory hypothesis) and comprehension (comprehension hypothesis). In addition, analogous to the findings that texts on multiple topics as well as highlighting can improve judgment accuracy (e.g., Gier et al., 2010; Prinz et al., 2020a), we assumed that headings would lead to greater metamemory (metamemory hypothesis) and metacomprehension accuracy (metacomprehension hypothesis) in terms of both bias and relative accuracy. Finally, we hypothesized that learners' cognitive load would differ depending on whether headings are present or not (e.g., Beege et al., 2020; Schneider et al., 2018). Specifically, we expected that headings would reduce extraneous load and increase germane load. We expected no variation in intrinsic load because the headings did not add information to the texts and therefore did not alter their complexity (cognitive-load hypothesis; cf. Mautone & Mayer, 2001).

2 | METHOD

2.1 | Sample and design

Based on a recent meta-analysis on signaling, which revealed a medium-to-large effect size for organizational signals within texts (e.g., headings or summaries) on learners' retention performance (Hedges's $g = 0.71$; Schneider et al., 2018), we expected to find medium-to-large effects of headings in our study. A respective power analysis yielded a required sample size of $N = 52$ ($\alpha = .05$, $\beta = .20$, $d = 0.70$; Faul et al., 2007). Therefore, 52 university students enrolled in educational science participated in this study. One participant from

the headings group had to be excluded because the time spent on studying the texts was too short to allow for careful reading (reading time in seconds relative to word count = 0.03). The mean age of the remaining 51 participants was 22.82 ($SD = 4.26$) years, and 82% were female (18% male, none non-binary). On average, the participants were in their 3.32 ($SD = 2.08$) semester of university studies. They could choose between course credit or monetary compensation for taking part in the study.

The experiment had a one-factorial between-subjects design with text type as the independent variable. Participants were randomly assigned to the headings condition ($n = 25$) or the plain-text condition ($n = 26$). In the headings condition, the experimental texts contained headings, whereas the texts in the plain-text condition contained no headings. Memory, comprehension, metamemory accuracy, metacomprehension accuracy, and cognitive load were the dependent variables.

2.2 | Materials and measures

2.2.1 | Texts

The five texts that the participants read in this study covered different topics from the natural sciences, namely, diatoms, el Niño, folate, r/K selection theory, and moors. The texts were between 482 and 505 words long (excluding headings). Each text was subdivided into five to six paragraphs. In the headings group, a heading preceded each paragraph. Different types of headings can be included in a text. For example, headings might consist of short statements or questions related to the topic to follow. Concerning the recall of facts, research has shown that it makes no difference whether headings are provided as topic statements or questions (Hartley & Trueman, 1985). However, questions might be beneficial when it comes to achieving deeper comprehension. This assumption is supported by research on adjunct questions in texts that participants have to read and answer during studying. Such questions have been found to support not only factual knowledge, but also and particularly higher-order processes such as knowledge application (e.g., Osman & Hannafin, 1994). Therefore, we included headings in the form of questions in this study. The question headings were printed in boldface on a separate line. The function of the headings was to foreshadow the upcoming topic. They did not convey any additional content information. Thus, the text versions with and without headings retained the same information. Figure 1 presents an example text paragraph for each condition. All five texts with headings as used in the present study are provided in the Supplementary Material.

2.2.2 | Prior knowledge

Participants were asked about their level of prior knowledge concerning the five text topics. They had to indicate their prior knowledge on a 5-point Likert scale ranging from 1 (*no prior knowledge*) to 5 (*very comprehensive prior knowledge*), respectively.

FIGURE 1 Example paragraph from the text on diatoms per condition.

Plain-text group	Headings group
At places in the sea where many diatoms accumulate, many crustaceans that feed on diatoms also gather. Their excretions bind together with dead algae and build a permanent stream of particles in the ocean, which is called marine snow. The marine snow sinks into deeper and darker zones and provides, for example, the soil organisms of the deep sea with food. When diatoms are not eaten but die, their shells are preserved and sink to the seabed together with the CO_2 they contain. This process, in which carbon is transported to the deep sea in the organisms' bodies and remains stored there over the longer term, is called biological pump. On the seabed, so-called diatomaceous earth is then formed, a thick layer of dead diatoms.	**Why are diatoms important even after their death?** At places in the sea where many diatoms accumulate, many crustaceans that feed on diatoms also gather. Their excretions bind together with dead algae and build a permanent stream of particles in the ocean, which is called marine snow. The marine snow sinks into deeper and darker zones and provides, for example, the soil organisms of the deep sea with food. When diatoms are not eaten but die, their shells are preserved and sink to the seabed together with the CO_2 they contain. This process, in which carbon is transported to the deep sea in the organisms' bodies and remains stored there over the longer term, is called biological pump. On the seabed, so-called diatomaceous earth is then formed, a thick layer of dead diatoms.

2.2.3 | Memory and comprehension

For each text, participants had to answer three detail questions, assessing memory, and three inference questions, assessing comprehension. Detail questions required the recall of specific information, such as terms, numbers, and characteristics. For example, a detail question on the text on diatoms was: "What is the thick layer of dead diatoms on the seabed called?" Inference questions required deeper understanding obtained by integrating information on such aspects as relationships, processes, and differences. For instance, an inference question on the text on diatoms was: "What happens with diatoms after they die?" Participants received 1 point for providing a correct answer and 0 points for providing no or an incorrect answer. Two raters independently scored 10% of the participants' answers, with high interrater agreement on the detail, Cohen's $\kappa = .97$, 95% CI [.92, 1.00], and inference questions, Cohen's $\kappa = .92$, 95% CI [.83, 1.00].

2.2.4 | Metamemory and metacomprehension judgments

After reading each text but before seeing any questions, the participants made their judgments. They judged the number of questions (0–3) in the detail category (metamemory judgments) and inference category (metacomprehension judgments) they would be able to answer correctly.

2.2.5 | Metamemory and metacomprehension accuracy

Metamemory and metacomprehension accuracy were operationalized in terms of bias and relative accuracy, respectively. Bias was calculated by subtracting each participant's actual number of correct answers from the participant's judged number of correct answers in each category for each text (e.g., Schraw, 2009). A positive value indicated overconfidence, a negative value underconfidence, and a value of zero a perfectly accurate judgment. Relative accuracy was computed via the intraindividual gamma correlation between a participant's judgments and actual performance scores across the texts. A larger positive correlation indicated greater relative accuracy because a participant expected better performance on texts where they actually performed better and expected worse performance on texts where they actually performed worse. A larger negative correlation indicated poorer relative accuracy because a participant expected worse performance on texts where they performed better and expected better performance on texts where they performed worse.

2.2.6 | Cognitive load

We assessed the cognitive load participants experienced during reading with the Cognitive Load Questionnaire by Klepsch et al. (2017). The questionnaire consists of eight items assessing intrinsic (e.g., "The texts were very complex"), extraneous (e.g., "While reading the texts, it was exhausting to find the important information"), and germane load (e.g., "I made an effort not only to understand several details, but to understand the overall context"). The items were slightly adapted to refer to the task of reading texts. The participants indicated their answers on a 5-point Likert scale ranging from 1 (*does not apply at all*) to 5 (*applies completely*). Internal consistency was rather low for the subscales on intrinsic ($\alpha = .50$) and germane ($\alpha = .38$) load but acceptable for the subscale on extraneous load ($\alpha = .73$).

2.2.7 | Perceptions of the headings

In an exploratory manner, we assessed the participants' perceptions of the headings. To do so, the headings group completed a questionnaire with eight statements concerning the headings (e.g., "The

TABLE 1 Descriptive statistics for the statements concerning the headings within the texts.

Statement	M	SD
1: The headings helped me get a better overview of the texts' content.	4.08	0.91
2: The headings helped me focus more strongly on the content of the texts.	3.76	0.88
3: The headings contributed to the fact that I read the texts more attentively.	3.72	0.98
4: The headings encouraged me to think more about the content when reading the texts.	3.80	0.87
5: The headings helped me acquire a better understanding of the texts.	3.88	0.93
6: The headings helped me assess my understanding of the texts.	3.32	1.11
7: The headings disrupted my reading flow.	1.56	0.77
8: I often skipped the headings.	2.08	1.22

headings contributed to the fact that I read the texts more attentively"; see also Table 1). The participants provided their answers on a 5-point Likert scale ranging from 1 (*does not apply at all*) to 5 (*applies completely*).

2.3 | Procedure

Due to the COVID-19 pandemic during the university winter term 2020–2021, the participants took part in the study online at home, which took approximately 75 min. We applied three procedures to ensure that participants completed the study appropriately and to ensure high data quality. First, before taking part in the experiment, we asked the participants to seek out a quiet place where they could work undisturbed and concentrate for 75 min. In addition, we asked them to refrain from using their mobile phone or engaging in other activities during their participation. Second, at the end of the study, the participants could report technical problems or other complications that might have arisen and prevented them from working undisturbed. This item showed that no major issues occurred. Third, we recorded participants' reading time. No overly long reading times that might be indicative of distractions occurred.

During the study, the participants first provided informed consent. Then, they indicated their prior knowledge concerning the text topics. Afterwards, they studied the five texts, with a reading time of 10 min per text. After reading each text, they provided their metamemory and metacomprehension judgments. To do so, they received definitions of detail and inference questions. However, they were given no information about the specific questions that would be asked. After reading and providing judgments for the last text, the

participants completed three detail and three inference questions for each text (30 questions in total). They were instructed to answer each question in as much detail as possible and to put down a question mark if they could not remember anything at all. Then, they completed the cognitive load questionnaire. The participants in the headings condition additionally completed the questions on their perception of the headings. Finally, the participants answered some demographic questions.

3 | RESULTS

An alpha level of .05 was used for all analyses. We report Cohen's *d* as an effect size measure. Values of 0.30, 0.50, and 0.80 represent thresholds for small, medium, and large effects, respectively (Cohen, 1988). In the analysis for each dependent variable, we looked at the average scores across texts. For interested readers, we additionally provide the results for each text separately in the Supplementary Material (see Table A1). Note that the pattern of results was quite similar for most individual texts. In cases where the assumption of homogenous variance was not fulfilled, we report corrected degrees of freedom.

3.1 | Preliminary analyses

Generally, the participants' prior knowledge was very low and did not significantly differ between the headings group ($M = 1.20$, $SD = 0.22$) and the plain-text group ($M = 1.21$, $SD = 0.27$), $t(49) = 0.11$, $p = .912$, $d = 0.03$. The reading time in seconds for each text was divided by the text's word count because, due to the headings applied in the headings condition, the texts in the two conditions were of different lengths. The time spent reading the texts relative to their word count did not significantly differ between the headings group ($M = 0.83$, $SD = 0.29$) and the plain-text group ($M = 0.84$, $SD = 0.38$), $t(46.71) = 0.02$, $p = .984$, $d = 0.01$. Furthermore, neither participants' prior knowledge nor reading time was significantly related to the dependent variables. Therefore, they were not used as covariates in the analyses.

3.2 | Results concerning the memory and comprehension hypotheses

Concerning the detail questions, there was no significant difference between the headings group ($M = 1.82$, $SD = 0.62$) and the plain-text group ($M = 1.50$, $SD = 0.60$), $t(49) = -1.86$, $p = .069$, $d = 0.52$, failing to support the memory hypothesis.

With regard to the inference questions, there was a statistically significant difference between the groups, $t(49) = -3.58$, $p < .001$, $d = 1.00$. The headings group scored higher ($M = 1.54$, $SD = 0.77$) than the plain-text group ($M = 0.82$, $SD = 0.66$), supporting the comprehension hypothesis.

3.3 | Results concerning the metamemory and metacomprehension judgments

Although our hypotheses concerned the impact of headings on the accuracy of participants' judgments, we first examined their effects on the judgments themselves. Concerning participants' metamemory judgments, there was no significant difference between the headings group ($M = 1.49$, $SD = 0.44$) and the plain-text group ($M = 1.44$, $SD = 0.45$), $t(49) = -0.40$, $p = .691$, $d = 0.11$.

Similarly, concerning participants' metacomprehension judgments, there was no significant difference between the headings group ($M = 1.44$, $SD = 0.55$) and the plain-text group ($M = 1.24$, $SD = 0.46$), $t(49) = -1.41$, $p = .164$, $d = 0.40$.

3.4 | Results concerning the metamemory and metacomprehension hypotheses

3.4.1 | Bias

Concerning metamemory bias, there was no significant difference between the groups, $t(49) = 1.51$, $p = .138$, $d = 0.42$, failing to support the metamemory hypothesis. The headings group showed underconfidence ($M = -0.33$, $SD = 0.61$) that significantly differed from zero, $t(24) = -2.69$, $p = .013$, $d = 0.54$, whereas the plain-text group revealed high judgment accuracy ($M = -0.06$, $SD = 0.65$) that did not significantly differ from zero, $t(25) = -0.48$, $p = .633$, $d = 0.10$.

Regarding metacomprehension bias, the difference between the groups was significant, $t(49) = 2.42$, $p = .020$, $d = 0.68$. In line with the metacomprehension hypothesis, the headings group provided quite accurate judgments ($M = -0.10$, $SD = 0.89$) that did not significantly differ from zero, $t(24) = -0.54$, $p = .593$, $d = 0.11$, whereas the plain-text group showed overconfidence ($M = 0.42$, $SD = 0.61$) that significantly differed from zero, $t(25) = 3.49$, $p = .002$, $d = 0.69$.

3.4.2 | Relative accuracy

Relative metamemory accuracy could not be computed for six participants due to a lack of variance in their judgments. Contrary to the metamemory hypothesis, there was no significant difference between the groups in relative metamemory accuracy, $t(43) = -0.44$, $p = .666$, $d = 0.13$. Neither in the headings group ($M = .20$, $SD = .78$) nor in the plain-text group ($M = .10$, $SD = 0.73$) did relative metamemory accuracy significantly differ from zero, $t(22) = 1.21$, $p = .239$, $d = 0.25$ and $t(21) = 0.64$, $p = .529$, $d = 0.14$.

Relative metacomprehension accuracy could not be computed for 10 participants due to a lack of variance in their performance scores or judgments. In contrast to the metacomprehension hypothesis, there was no significant difference between the groups in relative metacomprehension accuracy, $t(33.49) = -1.18$, $p = .248$, $d = 0.37$. However, relative metacomprehension accuracy in the headings group ($M = .60$, $SD = .57$) significantly differed from zero, $t(20) = 4.78$, $p < .001$,

$d = 1.04$, whereas in the plain-text group ($M = .33$, $SD = 0.83$) relative metacomprehension accuracy did not significantly differ from zero, $t(19) = 1.79$, $p = .089$, $d = 0.40$.

3.5 | Results concerning the cognitive-load hypothesis

As expected in the context of the cognitive-load hypothesis, the headings group ($M = 3.40$, $SD = 0.63$) and the plain-text group ($M = 3.65$, $SD = 0.60$) did not significantly differ concerning their experienced intrinsic load, $t(49) = 1.48$, $p = .145$, $d = 0.41$. Contrary to the cognitive-load hypothesis, there was no significant difference in the perceived extraneous load between the headings group ($M = 2.21$, $SD = 0.69$) and the plain-text group ($M = 2.60$, $SD = 0.78$), $t(49) = 1.88$, $p = .067$, $d = 0.53$. However, in line with the cognitive-load hypothesis, the headings group ($M = 3.89$, $SD = 0.53$) experienced a significantly higher germane load than the plain-text group ($M = 3.56$, $SD = 0.51$), $t(49) = -2.28$, $p = .027$, $d = 0.64$.

In an exploratory manner, we conducted further mediation analyses to investigate a potential mediating role of cognitive load in the effects of headings on the dependent variables. However, although the path coefficients were in the assumed directions, the indirect effects testing for mediation were not significant. The results are provided in the Supplementary Material (see Table A2 and Figure A1).

3.6 | Perceptions of the headings

Overall, the participants perceived the questions as supportive for their learning (see Table 1). Specifically, the headings were rated as very helpful for getting an overview of the texts' content. In addition, the headings were neither skipped during reading nor did they hamper participants' reading flow.

4 | DISCUSSION

In this study, we examined to what extent headings within texts support learners' memory and comprehension as well as their judgment accuracy concerning their memory and comprehension. In addition, we explored to what extent headings affect learners' cognitive load.

When headings were implemented in the texts, learners' comprehension was enhanced. This result is in line with prior research and supports the assumption that headings assist learners in interpreting a text's organization so that more resources can be devoted to integrating information with each other and activated prior knowledge (e.g., Hyönä & Lorch, 2004; Mautone & Mayer, 2001). Contrary to the effect on comprehension, the effect on memory was not significant. This result stands in contrast to previous studies that found effects of headings on learners' memory (e.g., Hartley & Trueman, 1985; Sanchez et al., 2001). A potential explanation for the discrepancy is that the headings enhanced transfer-appropriate processing, which

refers to the phenomenon that performance is promoted when the processes during encoding match those during testing (e.g., Franks et al., 2000; Morris et al., 1977). In our study, the headings were formulated in the form of questions that tapped important relations and processes reported in the texts. Therefore, the headings might have increased learners' focus on these aspects, helping them gain a deeper understanding, but might not have similarly increased their attention with regard to specific details compared to when no headings were present. In accordance with this explanation, Reynolds and colleagues (Reynolds et al., 1979; Reynolds & Anderson, 1982) found that adjunct questions provided along with a text supported learners in correctly answering questions of the same type in a test after reading. For example, in the study by Reynolds et al. (1979), during reading, learners periodically answered one of three types of questions—requiring a technical term, a proper name, or a number—or answered no questions. On the posttest, learners who had been questioned performed better on questions requiring information of the same type as the earlier questions but differing in the specific content. Questioned learners also spent more time reading text segments that presented information of the type addressed in the questions. For example, learners who received questions requiring technical terms as answers spent more time reading segments containing such terms. The findings suggest that learners selectively attend to and process textual information in the category identified as relevant through questions within texts. However, with regard to our study, it should be emphasized that although the headings only yielded a significant positive effect on comprehension, this enhancement did not occur at the expense of learners' recall. Descriptively, memory performance was superior to comprehension performance, and compared with plain text the headings also tended to support memory albeit to a lesser extent.

Furthermore, our study extends prior research by showing that headings can also affect the accuracy of judgments learners make regarding their learning. Specifically, headings led to reduced bias in metacomprehension judgments. Learners who read texts without headings tended to be overconfident concerning their comprehension, whereas learners who read texts with headings made more accurate judgments. This finding is in line with prior research showing that learners often overestimate their comprehension (e.g., Maki et al., 2005; Wiley et al., 2016), and additionally indicates that headings can reduce this tendency. Like texts with varying topics and highlighting (e.g., Gier et al., 2010; Prinz et al., 2020a), headings might encourage learners to more attentively monitor their comprehension and use more valid cues when providing judgments about their comprehension. It also seems plausible that, because headings emphasize relevant information and increase the salience of the topic structure, more cognitive resources can be devoted to monitoring.

Turning to the relative accuracy of metacomprehension judgments, the results indicated no significant difference between learners who read texts with headings and learners who read texts without headings. Nonetheless, relative metacomprehension accuracy was .33 and did not significantly differ from zero for learners who read texts without headings, but was .60 and significantly differed from zero for learners who read texts with headings. Compared to the former value, the latter value more clearly exceeds the average level of relative metacomprehension accuracy of .24 found in a recent meta-analysis (Prinz et al., 2020a). Therefore, although the difference between groups was not significant, it seems that headings have at least some potential to support learners in accurately judging which texts they have understood better and worse.

Metamemory bias did not significantly differ depending on whether the texts included headings or not. Learners who read texts without headings tended to provide accurate judgments, whereas learners who read texts with headings were on average underconfident. Concerning relative metamemory accuracy, average levels were low and did not differ between groups. As indicated with regard to the results on comprehension versus memory, the headings might have focused learners' processing, including their monitoring, more on important relations and processes reported in the texts and less on specific details (cf. Reynolds et al., 1979; Reynolds & Anderson, 1982). This would reflect transfer-appropriate monitoring, which, analogously to transfer-appropriate processing, reflects the phenomenon that higher congruency between the monitoring and testing conditions supports the accuracy of metacognitive monitoring (e.g., Dunlosky & Nelson, 1997; Griffin et al., 2019). In line with the transfer-appropriate monitoring account, research on the test-expectancy effect has shown that informing learners about the nature of the test questions to be encountered after reading, namely memory or comprehension questions, can increase their metamemory or metacomprehension accuracy, respectively (e.g., Griffin et al., 2019; Prinz et al., 2020b). This is because information about the kind of knowledge to be acquired and tested guides learners towards monitoring and selecting more appropriate cues when judging their respective knowledge (i.e., memory or comprehension). Similarly, the headings might have provided implicit information about the kind of knowledge to be acquired and tested—that is, knowledge about relations and processes. This might have encouraged the learners to align their monitoring and cue use to this knowledge, which increased the accuracy of their metacomprehension judgments. At the same time, due to transfer-appropriate monitoring, the learners did not monitor or to a lesser extent monitored their memory of details. As a result, they were rather unsure about their achieved memory performance and hence tended to provide more underconfident metamemory judgments. The finding that learners who read texts without headings were on average accurate concerning their memory might reflect that, without any information about the kind of knowledge to be achieved from reading, learners typically expect that they should remember specific textual information rather than draw inferences (see, e.g., Wiley et al., 2005). Consequently, they monitor and rest upon cues reflecting their memory rather than their deeper comprehension (e.g., Thiede et al. 2010; Griffin et al., 2019). Thus, without headings, as a default, the learners might have more closely monitored their memory and used respective cues, resulting in more accurate metamemory.

Our study also provides empirical evidence concerning the mechanisms underlying the effects of headings. Specifically, learners who read texts with headings perceived higher germane load during

studying than learners who read texts without headings. This outcome is in line with the suggestion that signaling might enhance germane load (Schneider et al., 2018). The increased germane load likely shows that the headings supported the learners in activating relevant prior knowledge and organizing and integrating the information. Accordingly, the learners' perceptions of the headings were very positive in that regard. For example, they felt that the headings supported their overview of the texts' content and encouraged them to think more about the content during reading. Furthermore, although the finding failed to reach the level of significance, it should be noted that compared with learners who read texts without headings, learners who read texts with headings also tended to experience lower extraneous load. This result is in line with prior research showing that texts with organizational or visual signals required less cognitive capacity and reduced learners' extraneous load (Beege et al., 2020; Britton et al., 1982). Similarly, the headings might have helped to highlight important content and the texts' organization, assisting the learners in representing its topic structure. Moreover, due to the reduced extraneous load, resources might have been freed up not only for organization and integration but also for monitoring activities. Nonetheless, due to the lack of significance, this interpretation must be treated with caution. The results on cognitive load underline that headings have the potential to do more than just assist learners in memorizing and retrieving information. Headings seem to enable deep processing as well as active monitoring. Yet, the mediation analyses examining whether cognitive load mediated the effects of headings on the dependent variables were not significant, even though the individual path coefficients were largely in the expected directions. A potential explanation for this finding is that additional processes play a role in the relationships. For example, enhanced germane load does not inevitably mean that learners effectively use the available resources to draw relevant inferences and monitor relevant content, which could have led to associations that were less strong than expected. Alternatively, the sample might have been too small to uncover mediation effects, or the way in which cognitive load was assessed might have contributed to the outcomes.

5 | LIMITATIONS AND FUTURE DIRECTIONS

We incorporated headings in the form of questions that highlighted the text topics to follow because we assumed that this would be particularly effective at supporting deep learning processes. However, headings can be implemented in a variety of ways (e.g., Lemarié et al., 2012). For example, headings can be formulated as statements instead of questions. Headings can also differ concerning their wordiness or degree of elaboration (e.g., "Air Flow" vs. "Air Flow: Air Moves Faster across Top of Wing"; Mautone & Mayer, 2001). Moreover, headings can differ regarding whether they target relations and processes or details and facts reported in a text. Rather than indicating text content, headings can also contain puns or describe the meta-function of a text (e.g., "Introduction"). These variations should be taken into account in further investigations to determine the extent to which the conclusions drawn from this study are generalizable to other types of headings. In addition, we explained the finding that the headings implemented in our study (i.e., addressing important relations and processes) supported learners' comprehension and meta-comprehension accuracy with reference to transfer-appropriate processing and monitoring (e.g., Dunlosky & Nelson, 1997; Morris et al., 1977). Experimentally contrasting headings that address relations and processes and headings that address details and facts could provide further support for the transfer-appropriate-processing and -monitoring accounts. Specifically, it would be interesting to see whether headings that address details and facts more strongly enhance learners' memory and metamemory accuracy compared to their comprehension and metacomprehension accuracy, and vice versa for headings that address relations and processes. Such an outcome would indicate that headings need to be adapted to the specific learning goal associated with reading, that is, memorizing specific textual information or gaining a deeper understanding of the text content.

Furthermore, we assumed that the headings supported the learners in more carefully monitoring their learning and using more valid judgment cues. However, we did not assess learners' online monitoring or cue use. For example, it is unclear to what extent learners who read texts with headings were indeed aware of their improved comprehension when making their metacomprehension judgments. Therefore, to provide further support for the underlying mechanisms, future research should apply process measures. For instance, think-aloud procedures could reveal the extent to which participants monitor their learning during reading and the information they use when making judgments. Moreover, learners usually use their monitoring as a basis for making regulatory decisions, which in turn affect their performance (e.g., Nelson & Narens, 1990; Rawson et al., 2011; Thiede et al., 2003). The present study focused on the monitoring aspect only. Further studies should examine whether more accurate judgments due to headings in fact lead to more favorable regulatory decisions, for example, about which texts to restudy, and thus to enhanced performance.

Finally, we assessed cognitive load using the questionnaire by Klepsch et al. (2017) because it differentiates among three types of load. Using this questionnaire allowed us to draw more specific conclusions concerning the impact of headings on learners' cognitive processing. We administered the questionnaire offline and assessed participants' cognitive load retrospectively. That is, participants had to indicate their cognitive load after finishing the study tasks. Future studies might assess cognitive load constantly, such as with eye-tracking measures, or repeatedly during the learning process, such as by administering self-reports directly after participants read each text or paragraph. Such measures would provide even more refined insights into how headings affect cognitive load during reading (cf., e.g., Seufert, 2018). That we measured cognitive load retrospectively also limits causal conclusions that can be drawn with regard to the impact of cognitive load on memory and comprehension and the accuracy with which they were judged, as was examined in the

mediation models. Specifically, participants had to provide their judgments and answer the test questions before reporting their cognitive load. Therefore, assessing cognitive load online would also have the benefit of allowing clearer conclusions concerning its impact on learning and judgment accuracy.

6 | CONCLUSION

Our study revealed that headings supported learners in comprehending texts and accurately judging their comprehension. Moreover, headings affected learners' cognitive load, in particular by enhancing learning-relevant germane load. Concisely, headings within texts represent a powerful way to support self-regulated learning. Therefore, authors of instructional texts should consider including headings. Likewise, instructors should select texts that contain headings or, alternatively, insert their own headings. Learners might also be able to facilitate their own learning by choosing texts with headings or generating headings themselves.

CONFLICT OF INTEREST STATEMENT
The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT
Availability of data: The data are available from the corresponding author on request.

ORCID
Anja Prinz-Weiß https://orcid.org/0000-0002-1097-3442

ENDNOTE
[1] A revised view of the cognitive load theory (Sweller et al., 2011) distinguishes only between intrinsic and extraneous load. In our study, we maintained the traditional view because we sought to investigate the subjective complexity of the learning task (reflected by intrinsic load), the impact of the instructional design (reflected by extraneous load), and the effort learners invest in learning (reflected by germane load; cf. Klepsch & Seufert, 2020).

REFERENCES
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

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

Hartley, J., & Trueman, M. (1985). A research strategy for text designers: The role of headings. *Instructional Science*, *14*(2), 99–155. https://link.springer.com/content/pdf/10.1007/BF00052394.pdf?pdf=button%20sticky

Beege, M., Nebel, S., Schneider, S., & Rey, G. D. (2020). The effect of signaling in dependence on the extraneous cognitive load in learning environments. *Cognitive Processing*, *22*(2), 209–225. https://doi.org/10.1007/s10339-020-01002-5

Britton, B. K., Glynn, S. M., Meyer, B. J., & Penland, M. J. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology*, *74*(1), 51–61. https://doi.org/10.1037/0022-0663.74.1.51

Cauchard, F., Eyrolle, H., Cellier, J.-M., & Hyönä, J. (2010). Visual signals vertically extend the perceptual span in searching a text: A gaze-contingent window study. *Discourse Processes*, *47*(8), 617–640. https://doi.org/10.1080/01638531003628791

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

De Bruin, A. B. H., Roelle, J., Carpenter, S. K., Baars, M., & EFG-MRE. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educational Psychology Review*, *32*(4), 903–915. https://doi.org/10.1007/s10648-020-09576-4

Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language*, *36*(1), 34–49. https://doi.org/10.1006/jmla.1996.2476

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.

Franks, J. J., Bilbrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer appropriate processing (TAP) and repetition priming. *Memory & Cognition*, *28*(7), 1140–1151. https://doi.org/10.3758/BF03211815

Gier, V. S., Herring, D., Hudnell, J., Montoya, J., & Kreiner, D. S. (2010). Active reading procedures for moderating the effects of poor highlighting. *Reading Psychology*, *31*(1), 69–81. https://doi.org/10.1080/02702710903397082

Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, *37*(7), 1001–1013. https://doi.org/10.3758/MC.37.7.1001

Griffin, T. D., Wiley, J., & Thiede, K. W. (2019). The effects of comprehension-test expectancies on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 1066–1092. https://doi.org/10.1037/xlm0000634

Holley, C. D., Dansereau, D. F., Evans, S. H., Collins, K. W., Brooks, L., & Larson, D. (1981). Utilizing intact and embedded headings as processing aids with nonnarrative text. *Contemporary Educational Psychology*, *6*(3), 227–236. https://doi.org/10.1016/0361-476X(81)90003-5

Hyönä, J., & Lorch, R. F. (2004). Effects of topic headings on text processing: Evidence from adult readers' eye fixation patterns. *Learning and Instruction*, *14*(2), 131–152. https://doi.org/10.1016/j.learninstruc.2004.01.001

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, *48*(1), 45–77. https://doi.org/10.1007/s11251-020-09502-9

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Lemarié, J., Lorch, R. F., & Péry-Woodley, M.-P. (2012). Understanding how headings influence text processing. *Discours*, *10*, 3–22. https://doi.org/10.4000/discours.8600

Lorch, R. F., Jr., & Lorch, E. P. (1996). Effects of headings on text recall and summarization. *Contemporary Educational Psychology*, 21(3), 261–278. https://doi.org/10.1006/ceps.1996.0022

Lorch, R. F., Lorch, E. P., Ritchey, K., McGovern, L., & Coleman, D. (2001). Effects of headings on text summarization. *Contemporary Educational Psychology*, 26(2), 171–191. https://doi.org/10.1006/ceps.1999.1037

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, 93(2), 377–389. https://doi.org/10.1037/0022-0663.93.2.377

McNamara, D. S., & Magliano, J. P. (2009). Toward a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, pp. 297–384). Elsevier.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–171). Academic Press.

Osman, M. E., & Hannafin, M. J. (1994). Effects of advance questioning and prior knowledge on science learning. *The Journal of Educational Research*, 88(1), 5–13. https://doi.org/10.1080/00220671.1994.9944829

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell Publishing Ltd.. https://doi.org/10.1002/9780470757642.ch13

Prinz, A., Golke, S., & Wittwer, J. (2020a). How accurately can learners discriminate their comprehension of texts? A comprehensive meta-analysis on relative metacomprehension accuracy and influencing factors. *Educational Research Review*, 31, 100358. https://doi.org/10.1016/j.edurev.2020.100358

Prinz, A., Golke, S., & Wittwer, J. (2020b). To what extent do situation-model-approach interventions improve relative metacomprehension accuracy? Meta-analytic insights. *Educational Psychology Review*, 32(4), 917–949. https://doi.org/10.1007/s10648-020-09558-6

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

Reynolds, R. E., & Anderson, R. C. (1982). Influence of questions on the allocation of attention during reading. *Journal of Educational Psychology*, 74(5), 623–632. https://doi.org/10.1037/0022-0663.74.5.623

Reynolds, R. E., Standiford, S. N., & Anderson, R. C. (1979). Distribution of reading time when questions are asked about a restricted category of text information. *Journal of Educational Psychology*, 71(2), 183–190. https://doi.org/10.1037/0022-0663.71.2.183

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17, 19–36. https://doi.org/10.1016/j.edurev.2015.12.003

Sanchez, R. P., Lorch, E. P., & Lorch, R. F. (2001). Effects of headings on text processing strategies. *Contemporary Educational Psychology*, 26(3), 418–428. https://doi.org/10.1006/ceps.2000.1056

Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educational Research Review*, 23, 1–24. https://doi.org/10.1016/j.edurev.2017.11.001

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

Seufert, T. (2018). The interplay between self-regulation in learning and cognitive load. *Educational Research Review*, 24, 116–129. https://doi.org/10.1016/j.edurev.2018.03.004

Spyridakis, J. H., & Standal, T. C. (1987). Signals in expository prose: Effects on reading comprehension. *Reading Research Quarterly*, 22(3), 285–298. https://doi.org/10.2307/747969

Surber, J. R., & Schroeder, M. (2007). Effect of prior domain knowledge and headings on processing of informative text. *Contemporary Educational Psychology*, 32(3), 485–498. https://doi.org/10.1016/j.cedpsych.2006.08.002

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.

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., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, 47(4), 331–362. https://doi.org/10.1080/01638530902959927

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology*, 132(4), 408–428. https://doi.org/10.3200/GENP.132.4.408-428

Wiley, J., Griffin, T. D., & Thiede, K. W. (2008). To understand your understanding one must understand what understanding means. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 817–822). Cognitive Science Society. https://pdfs.semanticscholar.org/6292/53f86294a75b66b6e9f296fc582c8cbbc75f.pdf

Wiley, J., Griffin, T. D., Jaeger, A. J., Jarosz, A. F., Cushen, P. J., & Thiede, K. W. (2016). Improving metacomprehension accuracy in an undergraduate course context. *Journal of Experimental Psychology: Applied*, 22(4), 393–405. https://doi.org/10.1037/xap0000096

Wilhite, S. C. (1986). The relationship of headings, questions, and locus of control to multiple-choice test performance. *Journal of Reading Behavior*, 18(1), 23–40. https://doi.org/10.1080/10862968609547554

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.