Effects of Text-Belief Consistency and Reading Task on the Strategic Validation of Multiple Texts

Accepted for publication in *European Journal of the Psychology of Education* (2015)

Johanna Maier and Tobias Richter

University of Kassel, Department of Psychology

Holländische Str. 36-38, 34109 Kassel, Germany

Corresponding author:

Johanna Maier

E-mail: johanna.maier@uni-kassel.de

Telephone: +49-561 804 3856

Fax: +49-561 804 3586

Acknowledgments

The research reported in this article was supported by the German Research Association (Deutsche Forschungsgemeinschaft, DFG, grant RI 1100/5-2 and grant RI 1100/5-3). We would like to thank Caroline Ahrendts, Geraldine Klaus, Lina Krüger and Sandra Opoka for their help in preparing stimulus materials and collecting data. In addition, we especially thank Veronika Schmitt for her assistance in transcribing and coding the think-aloud data.

The World Wide Web is one major source used in the educational context that laypersons attend to when searching for information about science-related topics. Getting web-based information on science-related topics falls in the domain of multiple text comprehension (Perfetti, Rouet, & Britt, 1999), which requires that readers comprehend information from each of the texts and integrate the contents of different, often contradictory texts into one coherent referential representation (Perfetti et al., 1999). In addition, when reading multiple texts conveying divergent perspectives on the same scientific problem, readers often agree more strongly with one argumentative position over others (Maier & Richter, 2013a), which leads to new challenges that are of importance for formal and informal learning. In detail, readers often fail to construct a complete referential representation when multiple perspectives and divergent argumentative stances on a scientific controversy are available, but prefer belief-consistent over belief-inconsistent information (*text-belief consistency effect*, e.g., Eagly & Chaiken, 1993; Maier & Richter, 2013a, 2014). As such, readers' prior beliefs have been shown to lead to a confirmation or congeniality bias in information selection (e.g., Nickerson, 1998; Hart et al., 2009) as readers often select belief-consistent information and avoid belief-inconsistent information when choosing between different texts (Knobloch-Westerwick & Meng, 2011). Belief-consistent information is also preferred over belief-inconsistent information in comprehension and memory for conflicting information (e.g., Chinn & Brewer, 1993; Johnson & Seifert, 1994; Kardash & Scholes, 1995; Levine & Murphy, 1943; Maier & Richter, 2013a, 2014; Ross, Lepper, & Hubbard, 1975; Wiley, 2005). For example, using a recall task after presenting pro- and anticommunist messages to readers who held either of the two views, Levine and

Murphy (1943) found that participants had a better recall for belief-consistent arguments. Similar, Wiley (2005) found that readers recalled more belief-consistent arguments that were presented in single texts about a controversial topic. Maier and Richter (2013a) also found a text-belief consistency effect in a multiple-text comprehension study. In their study, the results revealed stronger situation models for the belief-consistent texts (for a similar result, see Maier & Richter, 2014). In sum, these results suggest that prior beliefs strongly affect the selection, processing, and memory of information in such a way that belief-inconsistent information is at a disadvantage.

However, forming and justifying an opinion about a controversially debated science-related topic in learning should be accompanied by an unbiased and critical learning process, which requires comparing the validity and epistemic status of information from different viewpoints (epistemic cognition, Barzilai & Zohar, 2014; Richter, 2011). We subsume such learning strategies under the term *strategic validation*. In the present study, we examined the extent that an argument in contrast to a summary task fosters the use of validation strategies in multiple text comprehension for undergraduate students. The reading goal triggered by an argument task can only be attained by considering belief-inconsistent information during comprehension (Hart, Albarracin, Eagly, Brechan, Lindberg, & Merrill, 2009; McCrudden & Sparks, 2014). Strategy use and the allocation of cognitive resources to belief-consistent and belief-inconsistent texts during reading were assessed on-line, measured by reading times and concurrent think-aloud protocols. In addition, we examined the relationship of resource allocation and strategy use on readers' referential representation of the state of affairs described in the texts (situation model, Kintsch, 1988).

In the next section, we address the effectiveness of two different reading tasks typically investigated in multiple text comprehension. Afterwards, we discuss the reading strategies that are likely to be promoted by the reading tasks in more detail.

Reading Task and Multiple Text Comprehension

Reading is a goal-directed activity in which readers adjust their specific strategies in accordance with their reading goals (McCrudden & Schraw, 2007). In the goal-focusing model of relevance, McCrudden and colleagues (e.g., McCrudden, Magliano, & Schraw, 2010, 2011) assume that given and personal intentions influence reading goals, which in turn influence text processing. Text processing is then assumed to influence learning. Personal intentions are internally-generated standards a reader brings to an educational setting, whereas given intentions are externally-provided cues, for instance instructions given by a teacher. Both intentions act in concert to form reading goals, which help readers to focus their attention on the identification and processing of text information relevant to such goals.

Consistent with this proposal, ample empirical evidence has shown that reading tasks influence text-processing strategies (e.g., Hagen, Braasch, & Bråten, 2012; Lorch, Lorch, & Klusewitz, 1993; Narvaez, van den Broek, & Ruiz, 1999; van den Broek, Lorch, Linderholm, & Gustafson, 2001) and subsequent comprehension outcomes (e.g. Bråten & Strømsø, 2010; Britt & Sommer, 2004; Gil, Bråten, Vidal-Abarca, & Strømsø, 2010; Richter, 2003). One prominent distinction of reading tasks is between *argument* and *summary tasks* (e.g., Bråten & Strømsø, 2010; Gil et al., 2010; Naumann, Wechsung, & Krems, 2009; Stadtler, Scharrer, Skodzik, & Bromme, 2014; Richter, 2003; Wiley & Voss, 1999). Argument tasks require that readers develop a point of view about a scientific topic, whereas summary tasks require that readers memorize the factual information provided by the text. Given these different demands, it is reasonable that an argument task not only leads to more active and critical processing but as a result also to deeper and more integrated mental representation of (multiple) texts (Wiley, Goldman, Graesser, Sanchez, Ash, & Hemmerich, 2009).

The empirical evidence, however, regarding the effectiveness of argument tasks for adult readers is inconsistent. For example, a seminal study from Wiley and Voss (1999) found that reading to write an argumentation—in contrast to the task of writing a narrative, summary or explanation—was beneficial for multiple text comprehension as indicated by more causal and integrated essays produced by undergraduate university students. In the same vein, Stadtler and colleagues (2014) found that undergraduates in an argument condition acknowledged and reported existing conflicts and multiple perspectives on the discussed topic, whereas undergraduates in a summary condition provided only a one-sided situation of the scientific controversy and failed to report existing conflicts. Contrary to these findings, Gil and colleagues (2010) found that undergraduate students receiving an argument instruction performed worse on a verification task compared to readers in a summary condition. Bråten and Strømsø (2010) showed with an intertextual inference verification task that both a summary and an argument instruction lead to an equally integrated topic understanding for university students. However, participants in both conditions developed a more strongly integrated mental representation of the multiple texts compared to readers receiving a global understanding reading goal instruction.

The inconsistent empirical findings with regard to the effectiveness of argument tasks call for further research investigating reading comprehension strategies stimulated by an argument versus a summary task.

Effects of Reading Tasks on the Strategic Processing of Multiple Texts

Reading comprehension strategies can be viewed as “deliberate, goal-directed attempts to control and modify the reader’s efforts to decode text, understand words, and construct meanings of text” (Afflerbach, Pearson, & Paris, 2008, p. 368). However, because reading goals signal readers not only why and what but also how they should read (McCrudden & Sparks, 2014), argument and summary tasks are likely to prompt readers to use different types of reading strategies. An argument task should foster the strategic validation of information including cognitive activities that are aimed at validating the knowledge claims raised in multiple texts by checking the internal consistency of arguments and their consistency with other information in the same and different documents (Richter, 2003, 2011; Schroeder et al., 2008). In contrast, summary tasks necessitate reading strategies that are directed at remembering the central facts (i.e., memorization strategies) without the need to validate the epistemic status of information (Barzilai & Zohar, 2014). In line with this assumption, Richter & Schmid (2010; see also Barzilai & Zohar, 2014) found a strong positive effect of the goal to develop a justified point of view on cognitive activities that can be viewed as validation strategies: consistency checking and knowledge-based validation. In the study by Richter and Schmid (2010), the goal to memorize facts was unrelated to the use of consistency checking and was even negatively related to the use of knowledge-based validation.

To the best of our knowledge, only one study in the field of multiple text comprehension addressed reading comprehension strategies triggered by argument vs. summary tasks using note-taking as an indicator of strategy use (Hagen et al., 2012). Hagen and colleagues found no general differences in the use of reading strategies (paraphrases, intertextual and intratextual elaborations) between the argument and summary conditions. However, they found a positive relationship between intertextual elaboration and intra- as well as intertextual comprehension outcomes in the argument condition, but no such relationship in the summary condition. The results notwithstanding, note-taking might not be a valid indicator of strategy use, because it could stimulate strategies that are not activated during normal reading. Moreover, Hagen and colleagues did not assess prior beliefs, which can have a strong influence on the comprehension of multiple texts with conflicting information (e.g., Maier & Richter, 2013a).

Rationale of the Present Experiment

According to the taxonomy of goals provided by McCrudden and colleagues (2011, p.7), argument and summary tasks can serve as general-purpose oriented relevance instructions that are externally-provided, but the two reading tasks are likely to differ in the type of strategies they promote. Argument tasks should foster the strategic validation of information, whereas summary tasks should promote the use of memorization strategies. Following the goal-focusing model (McCrudden et al., 2011), we additionally assumed that personal (i.e., beliefs) and given (i.e., reading tasks) intentions in multiple text comprehension would affect the cognitive effort that readers invest in a given text in an interactive manner. In detail, readers will invest less cognitive effort for processing belief-inconsistent information when they follow a summary task, but rather focus on belief-consistent information in such a reading task condition (Richter, 2011; McCrudden et al., 2011; similar to selective exposure; see, for example Festinger, 1957). In contrast, in an argument task, which calls for deep processing of belief-inconsistent information, we expect readers to invest the cognitive effort needed for processing belief-inconsistent information. Accordingly, and given that reading times reflect cognitive effort, we expected longer reading times for the belief-consistent text in the summary condition, whereas this preference should be eliminated under an argument instruction (Hypothesis 1).

Furthermore, readers who spend less time reading a text will probably process this text in a more superficial manner. Given this assumption, readers following an argument task should engage in strategic validation processes as they strive to understand the full scope of a scientific controversy. Under this reading task, the processing and comprehension of belief-inconsistent information should be increased (Hypothesis 2). In contrast, readers following a summary task instruction should focus on memorization. Based on earlier research revealing readers' general preference for belief-consistent information in selection, processing, and comprehension (Hart et al., 2009; Maier & Richter, 2013a), we expected readers in the summary condition to pay more attention to belief-consistent information (Hypothesis 3).

Finally, we expected that the two effects caused by an argument task – i.e., spending more time reading belief-inconsistent information and engaging more strongly in strategic validation – should be beneficial for comprehension of the situation model for the belief-inconsistent text (Hypothesis 4).The predictions are summarized in Figure 1. Note that for conciseness, only the paths relevant for our assumptions are presented.

Method

Participants

Thirty-nine university undergraduates (27 women and 12 men) majoring in different social sciences (e.g., psychology or education) participated in the study. All participants were native German speakers and their average age was 25.5 years (*SD*=7.8). Participants were given the choice between a small monetary reward (8 Euros per hour) and course credit as reward for taking part in the study.

Text Material

Two texts arguing for contrary positions whether or not electromagnetic radiation from cell phones causes possible health risks were used as experimental material. One text argued that electromagnetic radiation emitted from cell phones causes health risks (*pro-stance*), whereas the other text argued that electromagnetic radiation causes no health risks (*contra-stance*). The selection of this topic was based on a pilot study with an independent sample of 38 university students who rated 25 topics for various criteria. In the pilot study, participants judged the topic of cell phones on a scale from 1=*not at all* to 6=*very much* as moderately interesting (*M*=3.03, *SD*=1.42), indicated that they possess only low prior knowledge (*M*=1.97, *SD*=1.08), and did not clearly prefer one argumentative stance in the controversy (agreement to pro-stance: *M*=3.16, *SD*=1.05, agreement to contra-stance: *M*=2.79, *SD*=0.96).

The texts were the same in writing style, argument number, and text structure. Readability (pro-text: 46, contra-text: 44, both moderate difficulty, determined with the German adaption of the Flesch's Reading Ease Index, Amstad, 1978) and length (pro-text: 1091words; contra-text: 1076 words) were highly similar. Both texts were written by the first author and were based on information from web appearances of reputable German magazines (e.g., Spiegel Online). The texts followed the same rhetorical structure (major claim - five supporting arguments - conclusion). An independent sample of 111 university students indicated in a pilot-study that the two texts were equally understood and that both were perceived to contain high-quality arguments with a clear stance towards the scientific issue (Online-Resource 1).

Comprehension Measure

Comprehension was assessed with a verification task (adapted from Schmalhofer & Glavanov, 1986), which has been used in earlier research to assess situation model strength in multiple text comprehension (e.g., Maier & Richter, 2014). Participants indicated whether or not test sentences contained information that matched the situation described in the texts. Eight sentences of three types of test items (paraphrases, inferences, distracters) were constructed for each text (48 test items in total; see Online-Resource 2 for examples). Paraphrase items were constructed by varying the word order of an original text sentence and replacing the key content words with synonyms. Inference items did not correspond to original text sentences, but readers needed to infer information presented in such sentences during the construction of a situation model. Finally, distracter items provided information loosely related to text content that was not explicitly stated in the text and could not be inferred from the texts.

Following the procedure proposed by Schmalhofer and Glavanov (1986), situation model strength for each text was based on the probit-transformed proportion of *yes* responses to inference items minus the probit-transformed proportion of *yes* responses to the distracter items.

The probit-transformation, in which the proportions are transformed to corresponding *z* values to which 5 is added to avoid negative values (e.g., Cohen, Cohen, West, & Aiken, 2003, p. 241), was used to normalize the distribution of the variables.

Reader Characteristics

Prior beliefs. Participants' prior beliefs about the scientific topic discussed in the texts were assessed with eight items on a scale ranging from 1=*totally disagree* to 6=*totally agree*. Four items assessed participants' agreement with the claim made by the pro-text (e.g.,: "I think that the electromagnetic radiation from cell phones and cordless phones is hazardous"; Cronbach's α=.82). Four items assessed participants' agreement with the contra-text stance (e.g.,: "I think that the electromagnetic radiation from cell phones and cordless phones is hazard-free": Cronbach's α=.79). Participants' (dis-)agreement to these two scales was used to define text-belief consistency (see *Results*).

Reading skills. To assess readers' reading skills, the sentence verification subtest of the German-speaking reading test ELVES (Cronbach's α=.87, Richter & van Holt, 2005) was used. In this task, participants judged the truth of simple assertions about abstract and concrete concepts. The test captures the efficiency of basic comprehension processes on the sentence level (i.e., lexical access, syntactic, and semantic integration) with combined test scores (accuracy and speed of a given response for each item; Richter & van Holt, 2005).

Procedure

Participants' prior beliefs were assessed one week prior to the experimental proper. In the experiment, participants were first made familiar with thinking aloud during reading by reading a short expository text. Participants were informed that they should continuously report all thoughts coming to their mind during reading. If participants remained silent for a minute, they were reminded to verbalize their thoughts. During this practice task, participants' thoughts were not recorded or coded. Afterwards, participants received the reading task instruction. Half of the participants were instructed to build a justified point of view whether or not electromagnetic radiation from cell phones is hazardous (*argument task instruction*). This instruction included the request to critically evaluate the text information and to judge the plausibility of the presented arguments. The other half of the participants were instructed to memorize as many facts as possible (*summary task instruction*). Afterwards, participants had 35 minutes to read the two texts in a web-based environment on a computer screen. Participants were able to choose the text they wanted to read first, to move freely back and forth within the texts, to switch between texts as often as they liked and to read the texts at their own speed. However, the order, in which the texts were presented on the website, was counterbalanced across participants. Participants' think-aloud comments and their reading time of the two texts were recorded while they were reading. To ensure that participants were aware of the reading task instruction during reading, participants received a prompt that matched their reading task every 7 minutes on average (the number of prompts was recorded). The prompt requested that participants critically evaluate the plausibility of an assumption made within the currently read text in the argument condition, and instructed participants in the summary condition to state the key facts from the last paragraph.

After reading, participants worked on the verification task in which they indicated by pressing one of two response keys (marked green=*yes* and red=*no*) whether or not a presented test sentence matched the situation that was described in the texts. Test items were presented sequentially in black letters on a white background (font type Arial bold, visual angle: 0.54 degrees) in random order. At the end of the experiment, participants were thanked and debriefed.

Protocol Analysis

Think-aloud protocols during reading are a valid method to examine strategic processes that readers engage in during reading (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). In the present study, the recordings of the think-aloud data were transcribed and segmented using comment(s) about the same core sentence or group of sentences as the unit of analysis (Coté et al., 1998). Think-aloud protocols were coded into three main categories (based on work from Afflerbach & Cho, 2009; Bråten & Strømsø, 2003; Coté, Goldman, & Saul, 1998; Crain-Thoreson, Lippman, & McClendon-Magnuson, 1997; Pressley & Afflerbach, 1995; Richter, 2003; Richter & Schmid, 2010; van den Broek et al., 2001; Weinstein, Husman, & Dierking, 2000; Zimmermann, 2000). The main categories were *strategic validation* that is assumed to be central for an argument task and *memorization* that is assumed to be central to a summary task. In addition, *organization and comprehension monitoring* was used as third main category. Comments that could not be coded into one of the categories were coded by a residual category. This category included general comments or associations that were not relevant to text comprehension (e.g., “I am done” or “OK”). Due to the fact that we were only interested in the influence of reading task on strategic validation and memorization, no details are provided for organizing and monitoring and for the residual category (information is available from the authors upon request). Online-Resource 3 provides a description and an example for each category.

Strategic validation. Comments that indicated the active use of prior knowledge and a strategic evaluation of the epistemic status of information were classified as strategic validation (Richter, 2003; Richter & Schmid, 2010). The first subcategory, *epistemic elaboration*, was coded when participants’ comments indicated the active use of prior knowledge to build inferences, to integrate currently read information into existing knowledge structures (i.e., generating personal experiences or everyday life examples), and to illustrate the texts’ content by creating mental images (cf. Van den Broek, Fletcher, & Risden, 1993; for an overview, see Caldwell & Leslie, 2010). Participants’ use of prior knowledge to judge the plausibility of information (Richter, 2011; Richter & Schmid, 2010) was also included. The second subcategory, *consistency checking within a document*, was coded when a participant validated text information in light of other information from the same text (partially overlapping but not identical to judgments of text coherence, Narvaez et al., 1999). Comments included searching a text for claim support or scrutinizing the argumentative and logical consistency of a text's argumentation (Richter, 2003; Richter & Schmid, 2010; see also Baker, 1989). In contrast, the third subcategory, *consistency checking across documents*, included validating the plausibility and consistency of information in light of information from the other text. This type of strategic processing is especially important for multiple text comprehension, because it requires scrutinizing the argumentative consistency between texts (intertextual relationships and inconsistencies; Maier & Richter, 2014). Finally, *evaluation of the text content* was coded when participants' comments included an evaluative judgment about the validity of text information without direct reference to prior knowledge or intertextual consistency.

Memorization. Memorization was coded when participants simply restated text information that indicated superficial text processing. In other words, memorization resembled a rehearsal of text information that did not elaborate beyond the content of the text (similar to text repetitions, Van den Broek et al., 2001; or paraphrases, Trabasso & Magliano, 1996).

In general, most units of analysis clearly fit into one category. However, on some occasions a verbal response was included in more than one category, for example, when participants first rephrased text information (i.e., paraphrasing) and in the same comment validated the information with regard to their prior knowledge (i.e., epistemic elaboration). The unit of analysis for these types of comments received two different codings. Parsing and coding of the think-aloud protocols were conducted by two independent coders that were blind to the experimental conditions. The inter-rater reliability of the coding system was good (Cohen's κ=.80), as determined by having two raters code 321 comments. Discrepancies in the ratings were resolved through discussion.

Design

The experimental design was a 2 (*text-belief consistency*: belief-consistent vs. belief-inconsistent, varied within subjects) x 2 (*reading task instruction*: argument vs. summary, varied between subjects) design. The order, in which participants read the texts (belief-consistent first vs. belief-inconsistent first, varied between subjects) was included as an additional factor. For the analysis of the reading times, reading skills and the number of prompts were included as covariates.

Results

Manipulation Check for Text-Belief Consistency

Participants' agreement to the two belief scales was used to define text-belief consistency. Overall, participants' mean agreement to the stance of the pro text (M=3.11, SD=1.00) was similar to participants' mean agreement to the stance of the contra text (M=3.04, SD=1.21), $t(38)$=0.31, p=.84. However, this does not mean that participants were fairly neutral with regard to the discussed issue. Rather, similar mean scores were due to the fact that approximately half of the participants (N=20) more strongly agreed with the stance of the contra text and that the other half more strongly agreed with the stance of the pro text (N=19). Hence, we determined text-belief consistency for each participant according to the argumentative position to which they more strongly agreed (as indicated by the agreement to the two belief scales). There was no participant that had the same scores on the two belief scales.

Exploratory Analyses of Strategy Use

Table 1 provides an overview of the mean frequency of each type of strategic processing as a function of reading task and text-belief consistency. Four subcategories served as indicators for strategic validation: 1) epistemic elaboration, 2) consistency checking within documents, 3) consistency checking across documents, and 4) evaluation of text content. Overall, participants made more comments coded as instances of epistemic elaboration (M=22.38, SD=9.5, range: 8-47) and consistency checking across documents (M=4.49, SD=4.98, range: 0-27). In contrast, only few comments were coded as consistency checking within documents (M=1.03, SD=1.29, range: 0-5) and evaluation of text content (M=1.71, SD=2.41, range: 0-11). Because all four subcategories were supposed to be indicators of strategic validation as broader construct and because of the low number of valid think-aloud comments in some subcategories, the four subcategories of strategic validation were combined into one summary score (Cronbach's alpha=.59). Think-aloud comments coded as memorization ranged from 0-56 (M=15.03, SD=12.71).

Given that (rare) count data is often positively skewed, which can lead to a non-linear relationship with other continuous variables (Cohen et al., 2003), we tested whether the frequencies of strategic validation and memorization were normally distributed. The Kolmogorov-Smirnov test for normality with Lilliefors-correction (Lilliefors, 1967; Massey, 1951) indicated that the distribution of strategic validation for the belief-consistent text (D=.16, p<.05) and memorization for the belief-inconsistent text (D=.16, p<.05) deviated significantly from a normal distribution. In contrast, the distributions of strategic validation for the belief-inconsistent text (D=.13, p=.08) and memorization for the belief-consistent text (D=.12, p=.17) did not deviate significantly from a normal distribution. Based on these results, we used a square root transformation to minimize the skewness in the count data as this transformation is commonly used for rare count data (Cohen et al, 2003, p. 245).

Effects of Text-Belief Consistency and Reading Task on Reading Times and the Think-aloud Measures

The hypotheses pertaining to effects of text-belief consistency and reading task instruction on reading times and think-aloud measures were tested with a split-plot analysis of variance (ANCOVA) with text-belief consistency as a within-subjects factor and reading task instruction as a between-subjects factor. The reading order was included as a control factor in all analyses. In the analysis of reading times as the dependent variable, reading skills and the number of prompts were included as covariates. All hypothesis tests were based on type-I error probability of .05. Descriptive statistics and intercorrelations of all variables are provided in Table 2. Based on the assumption of a medium effect size (f=.25 according to Cohen, 1988) and medium correlations (ρ=.5) between the levels of the independent variables in the population, the design and sample size of the experiment yielded a power (1-β) of .87 for detecting the focal interaction of text-belief consistency and reading task instruction (power was computed with the software G*Power 3; Faul, Erdfelder, Lang, & Buchner, 2007).

Effects on reading times. We found no main effects of text-belief consistency (F(1,33)=2.89, *n.s.*) and reading task instruction (F(1,33)=2.25, *n.s.*) on reading times. However, the analysis revealed a significant interaction of text-belief consistency and reading task instruction, F(1,33)=4.5?, p<.05, η_p^2=.12 (Figure 2). Based on mean differences, one-tailed paired-sample t tests were conducted to examine the interaction. In line with Hypothesis 1, participants in the argument condition spent an equal amount of time reading the belief-consistent (M=701.76, SE_M=30.83) and the belief-inconsistent text (M=723.64, SE_M=34.74), t(38)=-0.37, *n.s.* In contrast, participants in the summary condition read the belief-consistent text (M=822.99, SE_M=30.28) longer than the belief-inconsistent text (M=644.15, SE_M=34.12), t(38)=2.73, p<.05.

Effects on strategic validation. No main effects of text-belief consistency (F(1,35)=0.44, *n.s.*) and reading task instruction (F(1,35)=0.00, *n.s.*) were found in the analysis. However, we found a significant interaction of text-belief consistency and reading task instruction,

F(1,35)=4.78, p<.05, η_p^2=.12 (Figure 3a).[1] In line with Hypothesis 2, readers following an argument instruction used more validation strategies when reading the belief-inconsistent text (M=3.96, SE_M=0.25) compared to the belief-consistent text (M=3.48, SE_M=0.22), t(38)=-2.00, p<.05. Participants receiving the summary instruction used validation strategies for the belief-inconsistent text (M=3.59, SE_M=0.25) as often as for the belief-consistent text (M=3.85, SE_M=0.21), t(38)=1.08, *n.s.*

Effects on memorization strategies. The ANOVA on memorization strategies revealed a main effect of text-belief consistency, F(1,35)=7.16, p<.05, η_p^2=.17. Overall, readers used more memorization when reading the belief-consistent (M=2.58, SE_M=0.22) compared to the belief-inconsistent text (M=2.19, SE_M=0.22). No main effect of reading task instruction was found, F(1,35)=0.27, *n.s.*. However, we found an interaction of belief-consistency and reading task instruction, F(1,35)=9.33, p<.05, η_p^2=.21 (Figure 3b). In line with Hypothesis 3, readers receiving the argument instruction used memorization strategies equally often for belief-consistent (M=2.46, SE_M=0.32) and belief-inconsistent texts (M=2.52, SE_M=0.31), t(38)=-0.26, *n.s.* In contrast, readers receiving a summary instruction used less memorization when reading the belief-inconsistent (M=1.86, SE_M=0.30) compared to the belief-consistent text (M=2.69, SE_M=0.31), t(38)=4.20, p<.05.

Effects of Reading Times and Strategy Use on Situation Model Strength

In a final step, structural equation modeling (SEM) was used to examine whether or not the effects of reading task instruction and text-belief consistency on balanced resource allocation and the use of validation strategies and memorization together affect readers' comprehension of belief-consistent and belief-inconsistent texts. To test the hypothesized path-analytic model,

[1] For a better understanding, results based on raw frequencies are presented in Figure 3, whereas square-root transformed means (and their corresponding standard errors) are reported in the text.

analyses were conducted with the package lavaan (Rosseel, 2012) for the R environment for statistical computing (R Core Team, 2013). Given that the maximum likelihood estimation requires multivariate normality, which is often violated in small samples, we used maximum likelihood estimation in combination with the Bollen-Stine bootstrap to correct for non-normal data (Bollen & Stine, 1992; Rosseel, 2012; Schermelleh-Engel, Moosbrugger, & Müller, 2003). Moreover, the χ^2 test might fail to detect large deviances between the implied covariance matrix and the covariance matrix in small samples. Therefore, we also used the Standardized Root Mean Residual (SRMR) to evaluate the model fit (Hu & Bentler, 1999). A good model fit is indicated by SRMR values less than .08 (Hu & Bentler, 1999). The SRMR was used instead of the commonly reported Root Mean Square Error of Approximation (RMSEA), because in models with small sample sizes and few degrees of freedom, the RMSEA frequently exceeds cutoffs even for correctly specified models (Kenny, Kaniskan, & McCoach, 2014). In addition, we also investigated the Comparative Fit Index (CFI), which indicates a good model fit by values greater than .97 (Schermelleh-Engel et al., 2003). All statistical tests were assessed using a type-I error probability of .05.

The model was specified according to the goal-focusing model (McCrudden et al., 2010, 2010) and our assumptions and included only manifest (observed) variables. A direct path of reading task to balanced resource allocation and direct paths from balanced resource allocation and reading task to the use of strategic validation and memorization (separate for the two texts) were estimated. Direct paths from balanced resource allocation, reading task and the two types of strategies on the situation model for the two texts were estimated. Reading task (-1=summary task, 1=argument task) was contrast coded; all other variables were *z*-standardized. Strategic validation and memorization were allowed to covary.

The specified model had an excellent fit to the data, $\chi^2(4)=5.66$, $p=.28$, SRMR=0.07, CFI=.98 (Figure 4). As expected and in line with prior results, an argument task instruction was positively associated with balanced resource allocation for the two texts, $B=0.26$, $SE=.15$, $z=1.70$, $p<.05$, one-tailed. Furthermore, balanced resource allocation was positively associated with the use of validation strategies when reading the belief-inconsistent text, $B=0.56$, $SE=.13$, $z=4.23$, $p<.05$. Most interestingly, the situation model for the belief-inconsistent text was affected only by the use of validation strategies: A positive direct effect of validation strategy use emerged when reading the belief-inconsistent text, $B=0.32$, $SE=.16$, z=2.01, $p<.05$ but a negative direct effect of validation strategy use when reading the belief-consistent text, $B=-0.25$, $SE=.12$, $z=-1.99$, $p<.05$. In contrast, the situation model for the belief-consistent text was positively affected by memorization used when reading the belief-consistent text, $B=0.30$, $SE=.15$, $z=2.03$, $p<.05$ and negatively affected by validation strategies used when reading the belief-inconsistent text, $B=-0.42$, $SE=.19$, $z=-2.20$, $p<.05$.

Discussion

Readers' prior beliefs can bias the selection and comprehension of textual information in such a way that belief-consistent information is preferred (text-belief consistency effect, Eagly & Chaiken, 1993; Maier & Richter, 2013a). In line with the assumption that given reading tasks (i.e., argument and summary tasks) and readers' perspective (i.e., their prior beliefs) interactively influence processing of multiple texts (see McCrudden et al., 2010, 2011), we found differences in resource allocation and strategy use as a function of text-belief consistency and reading task. Readers following a summary task focused their resources on belief-consistent information as indicated by longer reading times and the use of more memorization strategies for the belief-consistent text. In addition, given that memorization strategies are inherent to a summary task, the positive link between the use of memorization strategies and the situation model strength for the belief-consistent text also (albeit indirectly) indicates that belief-consistent information receives more attention under a summary task. A summary task only requires a relatively superficial interaction with text and a more passive transfer of information (Wiley & Voss, 1999). Thus, these findings are consistent with the general notion that the text-belief consistency effect in the comprehension of (multiple) texts is connected to superficial processing (e.g., Hart et al., 2009; Knobloch-Westerwick & Meng, 2011; Maier & Richter, 2013a). Readers in the argument task condition, in contrast, allocated their processing resources in a more balanced way as they spent a similar amount of time reading both text types (belief-consistent vs. belief-inconsistent). Moreover, these readers used more validation strategies when reading the belief-inconsistent text. These findings are consonant with earlier research showing that an argument task calls for learning strategies that include the epistemic status of information (Barzilai & Zohar, 2014; Richter & Schmid, 2010). However, the findings extend earlier research insofar as they demonstrate that an argument task instruction increases strategic validation particularly for belief-inconsistent information. The reason for this latter finding probably is that an argument task leads to more argumentative thinking. Argumentative thinking entails that readers built and also defend their own point of view, which requires that readers are able to provide solid evidence for their position in a scientific controversy and also knowledge of counterclaims as part of an argumentative discourse (Kuhn, 1991). Hence, we think that an argument task fosters the need to put forward a constellation of propositions intended to justify one’s own standpoint and to refute the other side standpoint. In the case of comprehending belief-consistent and belief-inconsistent texts this requires an awareness of belief-inconsistent information to scrutinize the validity of this type of information.

In addition, the balanced resource allocation induced by an argument task instruction was positively associated with the use of strategic validation during reading the belief-inconsistent text, which improved the comprehension of the belief-inconsistent text but hindered the comprehension of the belief-consistent text. This finding was unexpected, because we assumed that the use of strategic validation overall should be beneficial for comprehension. Nevertheless, the result is in line with earlier research showing that a reduction of the text-belief consistency effect (i.e., similar strength of the situation models for the belief-consistent and the belief-inconsistent texts) comes at the cost of reduced situation model strength for the belief-consistent text (see Maier & Richter, 2013a). One possible explanation for the negative relationship between the use of strategic validation when reading one text type and the situation model of the other text type might be that the knowledge-based processes facilitated by an argument task benefitted only the situation model of the text that was read when these strategies occurred. Given that strategic validation is resource-demanding and effortful, readers might have been simply not able to maintain the same cognitive effort in processing the other text.

The present study advances the existing research in two ways. First, examining the combined effects of reading task and prior beliefs on strategy use during multiple text comprehension with the think-aloud methodology clarifies how and when argumentative tasks can lead to more active and critical processing. Moreover, the present results can help explain why previous studies investigating the effectiveness of argument vs. summary tasks found inconsistent results (e.g., Bråten & Strømsø, 2010; Gil et al., 2010; Wiley & Voss; 1999). Second, investigating comprehension outcomes with a verification task that included inference items allowed us to examine how and when learning strategies triggered by reading tasks are linked to comprehension on the level of the situation model. Based on our results, we suggest that the inconsistent empirical findings on the effectiveness of argument vs. summary task might be due to not assessing the moderating role of readers' prior beliefs. A summary task that prompts the use of memorization strategies can be beneficial only for the situation model of belief-consistent texts. In contrast, when an argument task prompts the use of validation strategies, these strategies only benefit the comprehension of belief-inconsistent texts.

Our results can be interpreted according to the process model of validation (Richter, 2011, 2015). The model assumes that epistemic monitoring processes are crucial for comprehension (e.g., Isberner & Richter, 2013, 2014; Richter, Schroeder, & Wöhrmann, 2009). In epistemic monitoring readers routinely use their prior knowledge and beliefs to validate the plausibility of text information as an integral part of text comprehension. Epistemic monitoring processes are linked to the construction of a situation model in such a way that information judged as implausible and belief-inconsistent is often not processed further, yielding plausibility biases and text-belief consistency effects (Maier & Richter, 2013a, 2013b; Schroeder et al., 2008). This model can also be used to interpret the preference for belief-consistent information found in the summary task condition of the present experiment. As a second assumption, the framework of validation suggests that readers can also engage in strategic and cognitively demanding epistemic elaboration processes directed at resolving the inconsistency (see Johnson-Laird, Girotto, & Legrenzi, 2004). The present results for the argument task condition fit well with this interpretation. That is, the use of validation strategies was promoted by an argument task and these strategies facilitated a deeper-level understanding of the belief-inconsistent text.

Even though the present study provides new insight into the processing and comprehension of multiple texts, we need to address several limitations. First, we assessed readers' situation model individually for each text type instead of assessing indicators of integration across texts (Strømsø, Bråten, & Britt, 2010). Hence, we focused on the comparability of the situation models that readers construct for each of the multiple texts. One reason for this focus was that each text's individual situation model can be viewed as the background for content integration (Britt, Perfetti, Sandak, & Rouet, 1999). Nevertheless, the integration of information across texts is one of the major challenges of multiple text comprehension (Perfetti et al., 1999) and further research should use additional tasks directed at capturing content integration (see Strømsø et al., 2010). A second limitation is that participants read texts about one particular scientific topic. Further research using different scientific topics is needed to show that the effects generalize across different topics. A final limitation is that think-aloud method used in this study warrants conclusions regarding slow and strategic comprehension processes but not regarding the role of fast and non-strategic processes (such as epistemic monitoring, Richter et al., 2009). Both types of processes act in concert during multiple text comprehension and are likely to account for the differences in processing belief-consistent and belief-inconsistent texts (Richter, 2015).

Reading tasks are an easy to use intervention in education and can direct and facilitate certain aspects of learning. For instance, in school and at universities, students are often required to memorize as much information as possible to pass an exam. The results of the present study suggest that such a reading task facilitates learning of belief-consistent information, but may hinder processing and comprehension of belief-inconsistent information. Especially in science education, where understanding a controversially debated science-related topic should include learning of information on both sides of the controversy, educators should reconsider the use of traditional assessment tasks evoking memorization strategies and a preference for belief-consistent information. Rather, educators should consider argument tasks as educational interventions that are able to lead to more argumentative thinking and stronger comprehension of belief-inconsistent information.

References

Afflerbach, P., & Cho, B. (2009). Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading. In S. Israel & G. Duffy (Eds.), *Handbook of reading comprehension research* (pp. 69-90). Mahwah, NJ: Erlbaum Associates.

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*, 364–373.

Amstad, T. (1978). *Wie verständlich sind unsere Zeitungen?* [How understandable are our newspapers?]. Unpublished doctoral dissertation, University of Zürich, Switzerland.

Baker, L. (1989). Metacognition, comprehension monitoring and the adult reader. *Educational Psychology Review, 1*, 3-38.

Barzilai, S., & Zohar, A. (2014). Reconsidering personal epistemology as metacognition: A multi-faceted approach to the analysis of epistemic thinking. *Educational Psychologist, 49*, 13-35.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods and Research, 21*, 205-229.

Bråten, I., & Strømsø, H.I. (2010). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes, 47*, 1-31.

Bråten, I., & Strømsø, H.I. (2003). A longitudinal think-aloud study of spontaneous Strategic processing during the reading of multiple expository texts. *Reading and Writing: An Interdisciplinary Journal*, *16*, 195-218

Britt, M.A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing task. *Reading Psychology*, *25*, 313 - 339.

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J.-F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209-233). Mawah, NJ: Erlbaum.

Caldwell, J., & Leslie, L. (2010). Thinking aloud in expository text: Processes and outcomes. *Journal of Literacy Research, 42*, 308-410.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science education. *Review of Educational Research*, *63*, 1-49.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mawah, NJ: Erlbaum.

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes, 25*, 1–53.

Crain-Thoreson, C., Lippman, M. Z., & McClendon-Magnuson, D. (1997). Windows of comprehension: Reading comprehension processes as revealed by two think-alouds. *Journal of Educational Psychology, 89*, 579–591.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt, Brace, Jovanovich.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* MIT Press: Cambridge, MA.

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*, 175-191.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row & Peterson.

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H.I. (2010). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology, 31*, 30-68.

Hagen, Å. M. M., Braasch, J. L. G., & Bråten, I. (2014). Relationships between spontaneous note-taking, self-reported strategies, and comprehension when reading multiple texts in different task conditions. *Journal of Research in Reading, 37*, 141-157.

Hart, W., Albarracin, D., Eagly, A. H., Brechan, I., Lindberg, M., Lee, K., & Merrill. L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*, 555-588.

Hu, L. , & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1-55.

Isberner, M.-B. & Richter, T. (2013). Can readers ignore implausibility? Evidence for nonstrategic monitoring of event-based plausibility in language comprehension. *Acta Psychologica, 142*, 15-22.

Isberner, M.-B. & Richter, T. (2014). Comprehension and validation: Separable stages of information processing? A case for epistemic monitoring in language comprehension. In D.N. Rapp & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 245-276). Boston, MA: MIT Press.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When discredited information in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1420-1436.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*, 640–661.

Kardash, C. A. M., & Scholes, R. J. (1995). Effects of pre-existing beliefs and repeated readings on belief change, comprehension, and recall of persuasive text. *Contemporary Educational Psychology, 20*, 201-221.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 1-22.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Knobloch-Westerwick, S., & Meng, J. (2011). Reinforcement of the political self through selective exposure to political messages. *Journal of Communication, 61*, 349-368.

Kuhn, D. (1991). *The skills of argument*. Cambridge, England: Cambridge University Press.

Levine, J.M., & Murphy, G. (1943). The learning and forgetting of controversial material. *Journal of Abnormal and Social Psychology, 38*, 507-517.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62*, 399-402.

Lorch, R. F., Lorch, E. P., & Klusewitz, M. A. (1993). College students' conditional knowledge about reading. *Journal of Educational Psychology*, 85, 239–252.

Maier, J., & Richter, T. (2013a). Text-belief consistency effects in the comprehension of multiple texts with conflicting information. *Cognition and Instruction*, *31*, 151-175.

Maier, J., & Richter, T. (2013b). How nonexperts understand conflicting information on social science issues: The role of perceived plausibility and reading goals. *Journal of Media Psychology, 25*, 14-26.

Maier, J. & Richter, T. (2014). Fostering multiple text comprehension: How metacognitive strategies and motivation moderate the text-belief consistency effect. *Metacognition & Learning, 9*, 54-71.

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for goodness of fit. *Journal of the American Statistical Association, 46*, 68-78.

McCrudden, M. T., Magliano, J., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals, and text processing: A mixed methods study. *Contemporary Educational Psychology, 35,* 229-241.

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2011). Toward an integrated view of relevance in text comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text Relevance and Learning from Text* (pp. 395-414). Greenwich, CT: Information Age Publishing.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19,* 113-139.

McCrudden, M. T., & Sparks, P. C. (2014). Exploring the effect of task instructions on topic beliefs and topic belief justifications: A mixed methods study. *Contemporary Educational Psychology, 39,* 1-11.

Naumann, A. B., Wechsung, I., & Krems, J. F.(2009). How to support learning from multiple hypertext sources. *Behavior Research Methods, 41*, 639-646.

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). Reading purpose, type of text and their influence on think-alouds and comprehension measures. *Journal of Educational Psychology, 91*, 488-496.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.

Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99-122). Mawah, NJ: Erlbaum.

Pressley M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Erlbaum: Hillsdale, NJ.

R Core Team (2013). *R: A Language and Environment for Statistical Computing* [Computer program]. Vienna: R Foundation for Statistical computing. Retrieved from http://www.R-project.org/

Richter, T. (2003). *Epistemologische Einschätzungen beim Textverstehen* [Epistemic validation in text comprehension]. Lengerich: Pabst.

Richter, T. (2011). Cognitive flexibility and epistemic validation in learning from multiple texts. In J. Elen, E. Stahl, R. Bromme, & G. Clarebout (Eds.), *Links between beliefs and cognitive flexibility* (pp. 125-140). Berlin: Springer.

Richter, T. (2015). Validation and comprehension of text information: Two sides of the same coin. *Discourse Processes, 52*, 337-352.

Richter, T., & Schmid, S. (2010). Epistemological beliefs and epistemic strategies in self-regulated learning. *Metacognition and Learning, 5*, 47-65.

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology, 96*, 538-598.

Richter, T., & van Holt, N. (2005). ELVES: Ein computergestütztes Diagnostikum zur Erfassung der Effizienz von Teilprozessen des Leseverstehens. *Diagnostica, 51*, 169-182.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*, 880-892.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48,* 1-36. Retrieved from http://www.jstatsoft.org

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23-74.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language, 25*, 279-294.

Schroeder, S., Richter, T., & Hoever, I. (2008). Getting a picture that is both accurate and stable: Situation models and epistemic validation. *Journal of Memory and Language, 59*, 237–259.

Stadtler, M., Scharrer, L., Skodzik, T., & Bromme, R. (2014). Comprehending multiple documents on scientific controversies: Effects of reading goals and signaling rhetorical relationships. *Discourse Processes*, *51,* 93-116.

Strømsø, H. I., Bråten, I., & Britt, M. A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction, 18*, 513-527.

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during reading. *Discourse Processes, 21*, 255–287.

van den Broek, P., Fletcher, C. R., & Risden, K. (1993). Investigations of inferential processes in reading: A theoretical and methodological integration. *Discourse Processes, 16*, 169–180.

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory and Cognition, 29*, 1081–1087.

Weinstein, C. E., Husman, J., & Dierking, D. R., (2000). Interventions with a focus on learning strategies. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 727-747) San Diego: Academic Press.

Wiley, J. (2005). A fair and balanced look at the news: What affects memory for controversial arguments? *Journal of Memory and Language, 53*, 95-109.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*, 301-311.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal, 46*, 1060-1106.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13-40). San Diego: Academic Press.

Table1: *Mean Use of each Type of Strategic Processing Varied by Text-Belief Consistency and Reading Task*

Strategy	**Category**	Argument task		Summary task	
		BC	BIC	BC	BIC
Strategic validation	Epistemic elaboration	10.0(5.6)	13.1(7.2)	11.5(4.8)	10.3(5.8)
	Consistency checking within documents	0.7(1.3)	0.5(1.0)	0.5(0.7)	0.4(0.8)
	Consistency checking across documents	2.1(2.8)	2.4(2.8)	2.4(3.7)	2.1(3.5)
	Evaluation of text content	0.3(0.5)	0.9(0.9)	1.2(1.8)	1.1(1.6)
Memorization	Memorization	7.6(5.4)	7.6(5.5)	9.2(8.6)	5.7(7.0)

Note. Standard deviations are presented in parentheses. BC=Belief-consistent Text, BIC=Belief-inconsistent Text.

Table2: *Descriptive Statistics and Intercorrelations of Independent Variables, Covariates, and Dependent Variables*

	M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1 Reading task	-0.02	1.01	1											
2 Text order	-0.03	1.01	.08	1										
3 Reading skills	18.50	4.61	-.18	.19	1									
4 Number of prompts	3.77	0.74	-.25	.03	.05	1								
5 Strategic validation(BC)	14.26	7.17	-.17	-.06	-.01	-.14	1							
6 Strategic validation(BIC)	15.36	8.42	.19	.01	.03	-.18	.47**	1						
7 Memorization(BC)	8.44	7.18	-.11	.11	.13	-.24	.44**	.43**	1					
8 Memorization(BIC)	6.59	6.33	.15	.08	.10	-.26	.26	.49**	.77**	1				
9 Reading times(BC)	760.22	178.75	-.43**	.21	.03	.61**	-.11	-.48**	-.18	-.36*	1			
10 Reading times(BIC)	695.66	191.88	-.04	-.08	-.03	.66**	-.01	.26	.03	.12	.01	1		
11 Situation model(BC)	1.59	0.88	-.20	.07	.23	.02	.03	-.22	.20	.10	.16	-.07	1	
12 Situation model(BIC)	1.79	0.73	.10	.31	.16	.10	-.09	.22	.06	.18	.08	.15	-.01	1

Note. BC=belief-consistent text, BIC=belief-inconsistent text. Reading task contrast coded: 1=argument vs. -1=summary. Text order contrast-coded: 1=belief-consistent first vs. -1=belief-inconsistent first. Reading skills: assessed with the sentence verification subtest of the ELVES (Richter & van Holt, 2005). Reading times: in seconds. Situation model: biased-corrected proportion of correctly inferred inference items.

* p<.05, ** p<.01 (two-tailed).

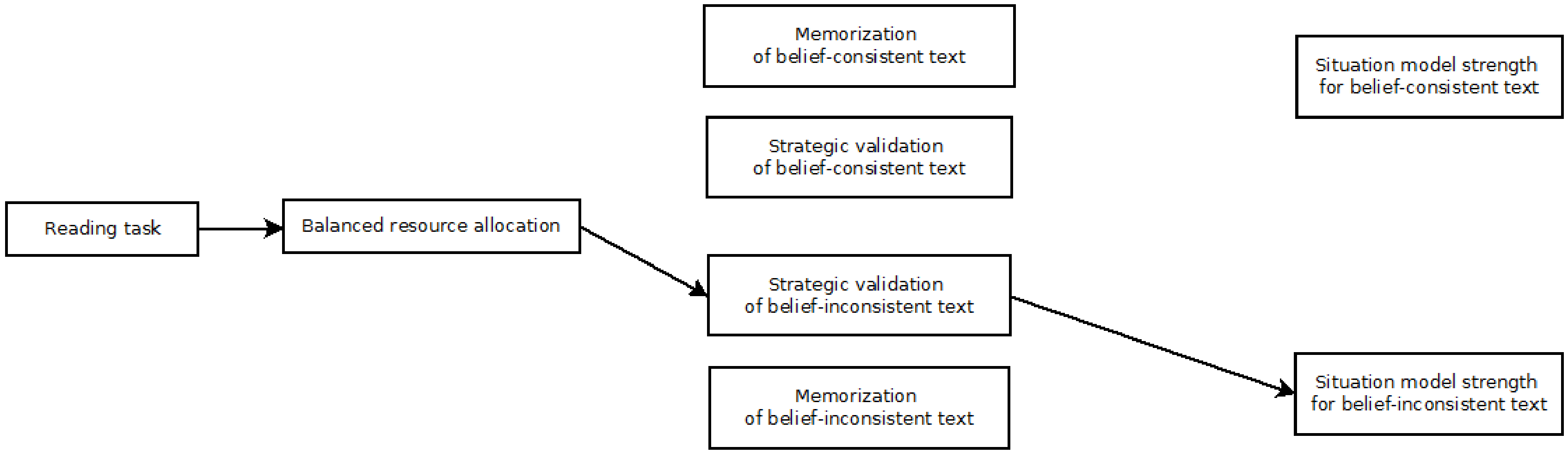

Figure1. Path diagram with expected relationships between reading task, resource allocation, strategy use and comprehension outcome.

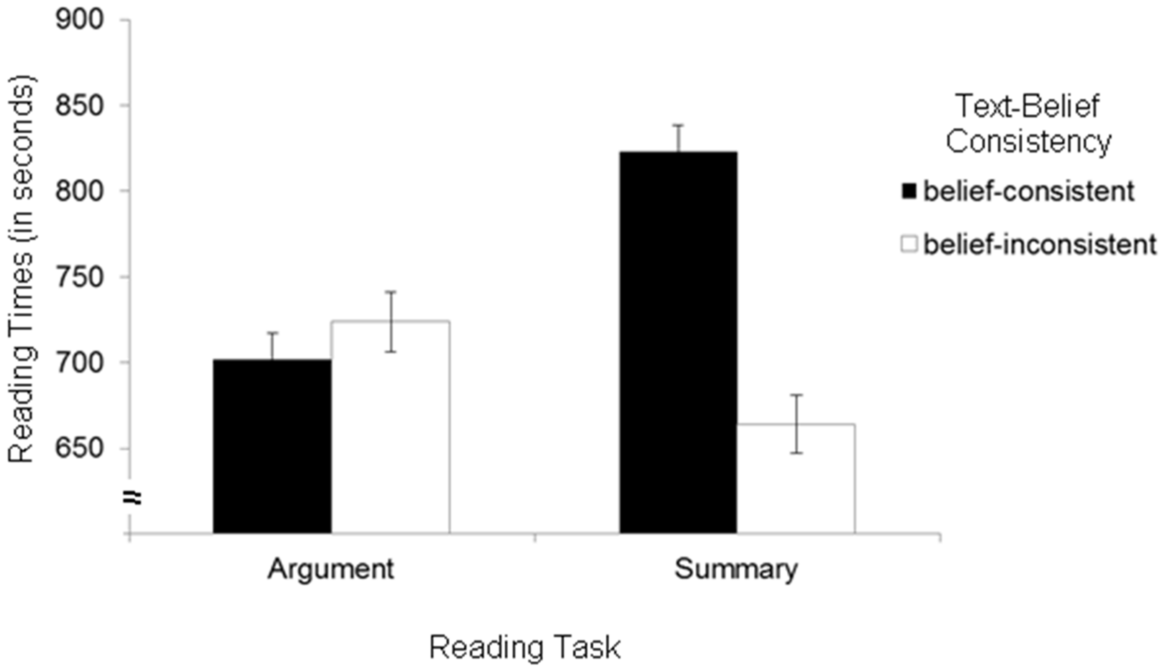

Figure2. Interaction of text-belief consistency and reading task (argument vs. summary) for reading times (error bars represent the standard error of the mean).

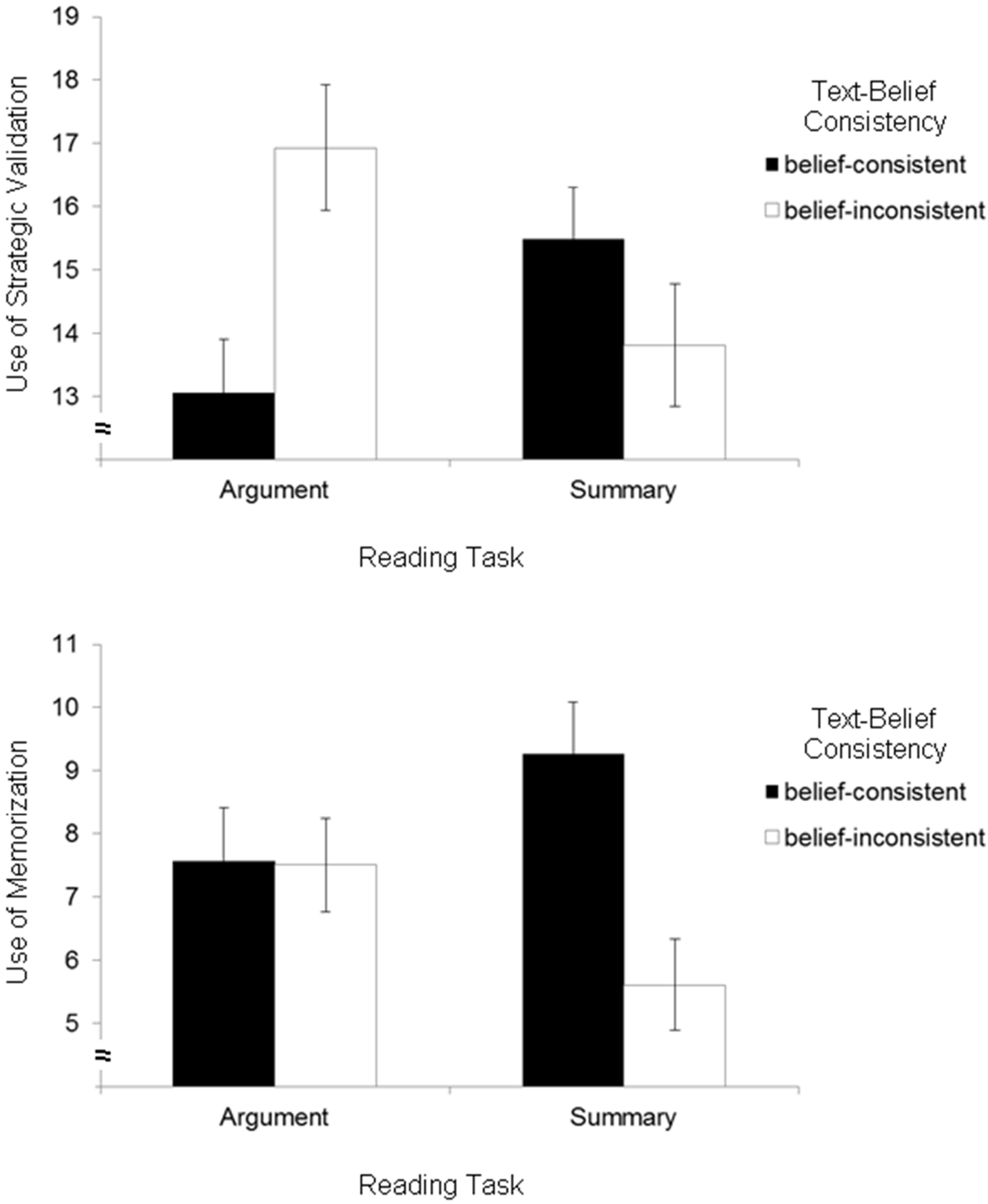

Figure3. Interaction of text-belief consistency with reading task (argument vs. summary) for a) use of strategic validation (raw data) and b) use of memorization (raw data), (error bars represent the standard error of the mean).

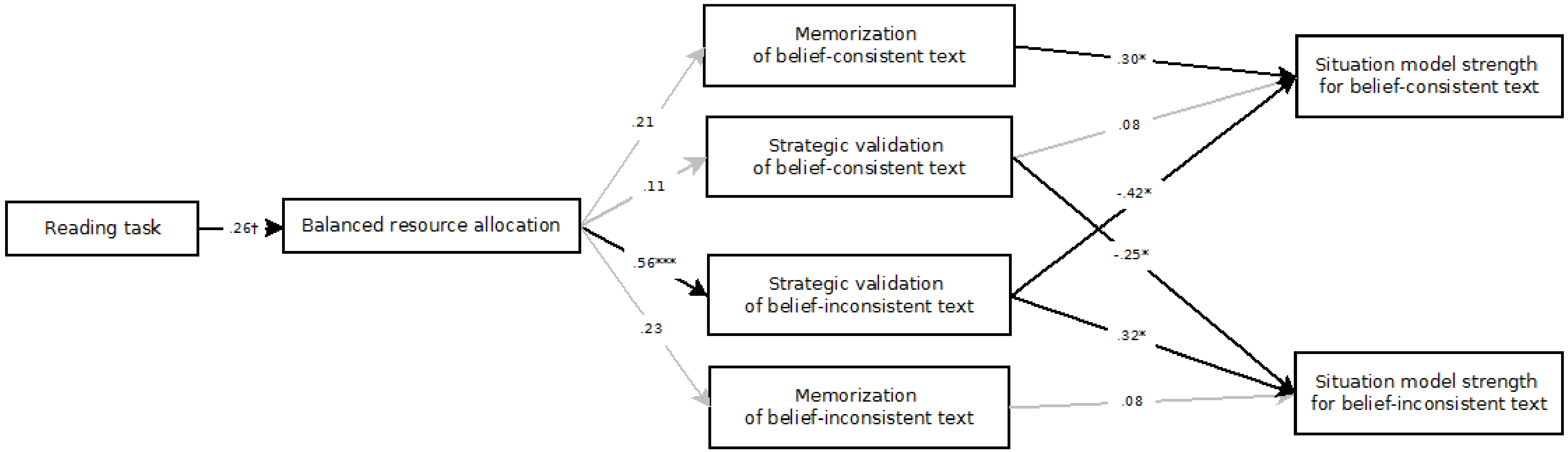

Figure4. Parameter estimates for the Structural Equation model. Reading task (-1=summary task, 1=argument task) was contrast coded, all other variables were z-standardized. Balanced resource allocation was indicated by the reading time difference of the belief-inconsistent and the belief-consistent text. For the sake of clarification, estimated paths from reading task on strategy use, as well as from balanced reading and reading task on situation model strengths are not visualized. Use of Memorization and strategic validation were allowed to covariate. Bold lines represent significant effects.

*** p<.001 ** p<.01, * p<.05 (two-tailed)

† p<.05 (one-tailed)