Psychology Press
Taylor & Francis Group

Metacomprehension judgements reflect the belief that diagrams improve learning from text

Michael J. Serra
Texas Tech University, Lubbock, TX, USA

John Dunlosky
Kent State University, Kent, OH, USA

In two experiments we systematically explored whether people consider the format of text materials when judging their text learning, and whether doing so might inappropriately bias their judgements. Participants studied either text with diagrams (multimedia) or text alone and made both per-paragraph judgements and global judgements of their text learning. In Experiment 1 they judged their learning to be better for text with diagrams than for text alone. In that study, however, test performance was greater for multimedia, so the judgements may reflect either a belief in the power of multimedia or on-line processing. Experiment 2 replicated this finding and also included a third group that read texts with pictures that did not improve text performance. Judgements made by this group were just as high as those made by participants who received the effective multimedia format. These results confirm the hypothesis that people's metacomprehension judgements can be influenced by their beliefs about text format. Over-reliance on this *multimedia* heuristic, however, might reduce judgement accuracy in situations where it is invalid.

Keywords: Multimedia learning; Metacomprehension; Diagrams; Metacognition; Heuristics.

Metacomprehension is a type of metacognition—thought about thought—that focuses on people's ability to evaluate their learning and comprehension of text materials. In most laboratory studies on metacomprehension, participants read text passages and then judge how likely they will be to correctly answer test questions about what they have read. Their learning of the materials is then tested. The accuracy of the metacomprehension judgements is estimated in numerous ways, such as by evaluating whether a particular factor (e.g., text difficulty) has a similar influence on judgement magnitude and on criterion test performance. Accuracy is important because the judgements play a functional role in the control of study (cf. Metcalfe & Finn, 2008) and such control is more effective when the judgements are more accurate (Thiede, Anderson, & Therriault, 2003). For this reason, much research has focused on explaining the low accuracy of metacomprehension judgements and exploring ways to make them more accurate (for reviews, see Dunlosky & Lipko, 2007, Lin & Zabrucky, 1998; Maki, 1998; Thiede, Griffin, Wiley, & Redford, 2009; Weaver, Bryant, & Burns, 1995).

One potential source of inaccuracy for metacomprehension judgements that has not received much empirical scrutiny is how people incorporate

Address correspondence to: Michael J. Serra, Department of Psychology, MS 2051, Psychology Building, Texas Tech University, Lubbock, Texas 79409-2051, USA. E-mail: michael.serra@ttu.edu

Experiment 1 was part of MJS's dissertation. We thank Katherine Rawson, William Merriman, Joel Hughes, Christopher Was, and Raymond Craig for serving on the dissertation committee, and for their helpful comments and suggestions regarding this research.

http://www.psypress.com/memory DOI:10.1080/09658211.2010.506441

their knowledge and beliefs about reading materials into their judgements (but see Cuevas, Fiore, & Oser, 2002; McCabe & Castel, 2008; Waddill & McDaniel, 1992). People's theories and beliefs about learning do not always reflect the actual influences of various factors on performance (Koriat, 1997), so basing judgements of text learning on them may introduce systematic biases into the judgements. For instance, people may believe that rereading improves text learning, and hence consistently judge that their text learning would be better after reading twice than after a single reading. This judgement would certainly reflect actual learning under some conditions, but it would inappropriately bias judgements under conditions where rereading does not improve learning (Callender & McDaniel, 2009; Rawson & Kintsch, 2005). Even if people's beliefs influence their judgements of text learning, these judgements may also be influenced by on-line monitoring of text processing, such as how easy a particular text was to read (e.g., Rawson & Dunlosky, 2002) or how much information was accessed about a text in the moments prior to making the judgement (Baker & Dunlosky, 2006; Morris, 1990). Thus an issue arises as to whether metacomprehension judgements will consistently be influenced by people's beliefs or whether on-line monitoring will reduce the influence of beliefs on the judgements.

To address this issue, in the present research we examined whether people's judgements would reflect the typical multimedia effect in which adding diagrams to text produces better learning than does text alone (for reviews, see Mayer, 2005; 2009). Importantly, judgements of text learning were collected in the present studies before, during, and after study, so that we could evaluate whether any beliefs about multimedia effects would influence judgements made prior to study, and whether this influence would persist even when on-line processes could influence the judgements (i.e., for judgements made during or after study). As discussed in detail later, this particular comparison among judgements was especially critical in Experiment 2 in which a multimedia format did not boost performance. Before describing the present experiments in detail, however, we will first consider how people's beliefs about learning may influence metacognitive judgements in general.

THE INFLUENCE OF BELIEFS ON METACOGNITIVE JUDGEMENTS

As proposed by Koriat (1997), one type of information that people consult when making metacognitive judgements for a task is their beliefs about their ability to perform the task, as well as the conditions under which the task is being performed (cf. Flavell, 1979; Hacker, 1998). Beliefs that affect memory judgements can take the form of self-referential beliefs about one's ability to remember information (i.e., ''I have poor memory for names'') or about the specific conditions of the learning situation (i.e., ''I'll do better on a recognition test than on a recall test'', Thiede, 1996). These two types of belief can be relatively independent of each other, even if they are about the same aspect of cognition. For example, Hertzog and colleagues (Hertzog, Dixon, & Hultsch, 1990; Hertzog, Dixon, Schulenberg, & Hultsch, 1987) have demonstrated that adults' memory self-efficacy is distinct from their beliefs about how memory functions in general.

Given that people's beliefs about cognition can be incorrect, using such beliefs to make metacognitive judgements may lead to judgement errors (cf. Nisbett & Wilson, 1977). Such errors arise when people evaluate their memory for simple materials (Benjamin, Bjork, & Schwartz, 1998; Koriat & Bjork, 2006; Serra & Metcalfe, 2009). For example, some paired associates seem easy to remember at the time they are studied, but will actually be difficult to remember during a later test. Koriat and Bjork (2006) described this error as *foresight bias*, and they created paired associates that consistently produced it. The words in these pairs had a strong backwards association but a weak forward association; for the pair ''rain – umbrella'', people are likely to generate ''rain'' when shown ''umbrella'' (strong backward association) but unlikely to generate ''umbrella'' when shown ''rain'' (weak forward association). When such pairs were studied and judged in a typical metamemory procedure, they were perceived as highly related and hence received high judgements of learning. At test, however, the stimulus word (rain) did not often produce the correct response (umbrella) because of the weak forward association (Koriat & Bjork, 2006). Thus beliefs about relatedness dominated the judgements even when relatedness of the words was not predictive of actual levels of performance. Similarly, Rhodes and Castel (2008) presented learners with to-be-remembered words in different font sizes. Larger words were easier to process than were smaller words, so learners consistently judged their memory to be better for the larger words than for the smaller words, despite the fact that font size had little effect on memory for the items.

Such errors might arise with text materials as well. For example, McCabe and Castel (2008) had learners read fictitious articles on cognitive-neuroscience research that either presented the supporting data as brain images (e.g., fMRI images) or as either bar graphs or topographical brain-activity maps (or the data were not presented visually at all). Despite the fact that all versions of the articles contained the same information, participants consistently rated the articles as more scientifically sound when they were accompanied by the brain images than in any of the other conditions, including when the articles were accompanied by the same data presented in either bar graphs or topographical brain maps (McCabe & Castel, 2008). Presumably, the readers' perception of the quality of the article was biased by the format of the data plot, not by any differences in the quality of either display. Although it would be ideal to know that the different formats produced the same level of comprehension (which was not measured), it nevertheless supports the main idea of the present study—that people's beliefs about multimedia effects may dominate their judgements, even when multimedia does not enhance learning.

It is worth noting that few metacomprehension studies have evaluated whether beliefs about reading materials are reflected in the magnitude of metacomprehension judgements. The studies available have focused on the interaction between people's background knowledge and the content of the materials, such as whether participants' familiarity with a domain is related to their judgements across more versus less familiar content areas (e.g., Glenberg & Epstein, 1987; Glenberg, Sanocki, Epstein, & Morris, 1987; Jee, Wiley, & Griffin, 2006; Maki, 1998; Moore, Lin-Agler, & Zabrucky, 2005). Consider the paper by Glenberg et al. (1987), who examined the role of participants' domain familiarity (i.e., their knowledge of various subject areas) on their metacomprehension judgements. Participants read texts on a number of topics and predicted their test performance for each one. They then answered questions that tested their learning for the texts. Participants used their familiarity with the topic of each text as the basis for their metacomprehension judgements rather than basing their judgements on their understanding of each particular text. Glenberg et al. (1987) described the way that participants used their domain familiarity to inform their metacomprehension judgements as ''the misapplication of a domain familiarity strategy'' (p. 132). Presumably this strategy was based on the participants' belief that they would perform better on questions over more-familiar topics than over less-familiar topics. As noted by Glenberg et al. (1987), although this might sometimes be a useful heuristic (i.e., when topic familiarity is related to comprehension), by relying on this strategy the participants ignored how well they understood the *particular* texts they had read. To foreshadow, we examined whether a similar bias would occur in Experiment 2, which included a multimedia format that did not boost performance.

OVERVIEW OF THE PRESENT EXPERIMENTS

Based on this evidence, we expected that participant's beliefs about text materials (i.e., that multimedia improves learning over reading text alone) would be reflected in their metacomprehension judgements. Nevertheless, beliefs about text format may not always be incorporated into metacomprehension judgements. In particular, on-line monitoring may reduce the impact of people's beliefs about material format on their judgement magnitudes. A student may initially believe that test performance will be better when reading a set of texts with diagrams than when reading texts alone. When actually reading the texts, however, the student may realise that performance will likely not be much better for the texts with diagrams—perhaps because the specific diagrams were not effective at aiding comprehension. Put differently, a student may believe that—in general—multimedia formats are better than are text alone, but while reading a particular set of texts, they may realise that the texts with diagrams are not going to support superior performance. In this case, their beliefs may have little—if any—influence on the magnitude of their judgements.

To explore these issues, we evaluated whether college students' beliefs about the instructional efficacy of the format of text materials—text with diagrams versus text alone—would be reflected in the magnitude of their judgements. We examined this relationship between beliefs and judgement magnitude when the format of the text materials influenced performance as expected (Experiments 1 and 2) and under a condition where the materials did not influence performance (Experiment 2). We chose multimedia materials because students prefer reading text with diagrams to reading text alone (Mayer & Massa, 2003) and the former often produce better learning (for reviews see Mayer, 2005, 2009). Note that two other studies have investigated the effects of multimedia on people's judgements of text learning. One reported that judgements were slightly higher for multimedia materials than for text alone, although this trend was not statistically significant (Waddill & McDaniel, 1992). In the other relevant study, by Cuevas et al. (2002), participants read either multimedia materials or text alone and made a post-study global judgement after they had finished reading the texts. Cuevas et al. found that between-participant correlations between global judgements and test performance were greater for a multimedia group than for a text-alone group, but this small advantage was only obtained in a subset of participants with low verbal ability. Most important, Cuevas et al. (2002) did not report judgement magnitudes, so the relationship between multimedia learning and judgements of text learning requires further scrutiny.

Our two main goals were to evaluate (a) whether people actually hold the belief that text with diagrams produces better learning than does text alone, and most important, (b) whether this belief is reflected in their metacomprehension judgements. For the former goal, we introduce the *beliefs about multimedia learning* (BAML) questionnaire, which directly assesses students' beliefs about the efficacy of multimedia formats. For the latter goal, students made pre-study global judgements, per-paragraph judgements, and post-study global judgements about their learning. When made *before* reading, judgements tap beliefs without any influence of reading and on-line monitoring per se. By contrast, per-paragraph and post-study global judgements may be largely influenced by on-line processing of the text (e.g., Rawson & Dunlosky, 2002) and hence may be less sensitive to beliefs about text format (cf. Hertzog et al., 2009). Thus we first expected that global judgements would be higher for multimedia materials than for text alone; as important, if people's beliefs about material format are moderated by on-line processing, then this effect will be smaller (or non-existent) for per-paragraph judgements. Note, however, this prediction only holds when multimedia materials do not produce better performance (as in the present Experiment 2), because it is under this condition that monitoring would potentially provide contradictory information concerning people's a priori beliefs about text format.

EXPERIMENT 1

Method

Participants, design, and materials. A total of 80 undergraduates from Kent State University participated in this study to partially fulfil a course requirement. The format of the materials for study (text with diagrams or text alone) was the only between-participants independent variable. Participants were randomly assigned to either study condition when they arrived at the laboratory, with the restriction that an equal number (40) were eventually assigned to the two groups. Regardless of their study condition, all participants later answered the same set of short-answer questions as a test of their understanding of the study materials.

Study and test materials. The study materials used in this experiment were adapted from Mayer's frequently used text with diagrams that explain how lightning storms develop (Mayer, 2009; Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Mayer, Steinhoff, Bower, & Mars, 1995; Moreno & Valdez, 2005) and were identical to the version used by Serra (2010). (Figure A1 in Appendix A shows an example paragraph, diagram, and two test questions used in Experiments 1 and 2, as well as an example photo used in Experiment 2.) More specifically, the materials consisted of a text passage of approximately 500 words describing in a cause-and-effect manner how lightning storms develop, and six explanatory diagrams. The text had six text paragraphs that were used in both the multimedia and text-only groups. In the multimedia group, a corresponding full-colour diagram accompanied each paragraph. The diagrams did not add any additional information to the paragraphs because they depicted information already provided in the text paragraphs. Understanding for each paragraph was tested with 2 short-answer questions, yielding a total of 12 questions.

Survey measures. To determine if participants believe that multimedia presentations produce better learning than does text alone, we created the beliefs about multimedia learning (BAML) questionnaire for them to answer at the outset of the experiment (see Table A1 of Appendix A). The purpose of these questions was to explicitly evaluate which of the two types of presentation (multimedia or text-only) participants believe produces better learning. The questions included in the BAML measured participants' beliefs about the efficacy of multimedia both for learners in general and for the individual. These questions were mixed with the questions from some of the learning-preference scales used by Mayer and Massa (2003) to make the BAML questions seem less conspicuous to the participants. The other scales included were the Verbal-Spatial Ability Rating (VSAR; a self-rating of verbal-spatial abilities from strongly verbal to strongly visual), the Santa Barbara Learning Style Questionnaire (SBCSQ; a self-rating of verbal-spatial abilities and preferences from strongly verbal to strongly visual), the Verbal-Visual Learning Style Rating (VVLSR; a self-rating of verbal-spatial abilities from strongly verbal to strongly visual), and the Learning Scenario Questionnaire (LSQ; a self-rating of verbal-spatial preferences from strongly verbal to strongly visual). For interested readers, these scales are presented in their entirety in Mayer and Massa (2003). Outcomes related to these scales were not directly relevant to the present study's goals and so will be reported elsewhere.

Procedure. The procedure of Experiment 1 was the same for the two groups except that a diagram accompanied each paragraph in the multimedia condition. As many as four participants at a time participated on individual PCs in the same room. These computers displayed the materials and questions to the participants and recorded all responses and latencies. At the onset of the experiment, the participants completed the learning-preferences and BAML questionnaires on the computer. The questions from the five questionnaires were presented in a random order to all of the participants and answered one at a time. Participants then read a general description of the task they would be performing and the type of questions they would be asked. Specifically, all participants were given the example of reading a text about how the brakes of a car work; the idea for this example was taken from another multimedia text frequently used by Mayer and his associates (see Mayer, 2009). The instructions did not mention diagrams for *either* condition, nor was any text (or diagrams) actually shown. Given that this hypothetical text was described as one that might appear in a science textbook, it was assumed that the participants, who were all college students, would be familiar with this type of material. The short-answer question, "What happens when the brake shoes press against the drum?" (Mayer, 2009), was also included in the instructions so that the participants knew what type of questions they could expect. An example of an acceptable correct answer, "The drum and wheel stop or slow down", was also provided.

The description of the task was slightly different for the two groups in that the participants knew whether they would be reading either a text or a text with diagrams (but they did not know that other participants might be viewing a different version of the materials, nor was there mention of other possible versions of the materials). Specifically, the instructions described the critical study materials as either "a short science text about how lightning storms develop" (text-only group) or "a short science text with diagrams about how lightning storms develop" (multimedia group). Participants had to know the format of the materials they would be studying so that they could make the pre-study global judgements appropriately. The instructions also included detailed explanations of the judgements the participants would be making.

After reading the instructions, participants made a pre-study global judgement; they estimated how well they felt they would understand how lightning storms develop (on a scale from 0 to 100) when they finished studying it. Participants then read the six paragraphs one at a time on the computer in serial order. Only one paragraph was visible onscreen at a time. In the multimedia group a corresponding diagram accompanied each paragraph. Participants were instructed to study each paragraph until they felt they understood it well enough to be able to answer questions on it later. By clicking a "Finished Studying" icon on the screen, participants indicated that they had finished studying that paragraph and were ready for the

TABLE 1
Percentage of participants endorsing beliefs about multimedia learning

	Experiment #	
Belief	*1*	*2*
Generalised beliefs		
The selected type of presentation would produce the best learning		
Diagrams only	1.3	0.8
Mostly diagrams with little text	25.0	13.3
Equal-parts text and diagrams	46.3	70.8
Mostly text with few diagrams	26.3	15.0
Text only	1.3	0.0
The selected section of a textbook chapter would produce better learning		
A section with diagrams	93.8	92.5
A section without diagrams	6.3	7.5
Belief about the self		
The selected textbook chapter would produce better learning for myself		
A textbook chapter with diagrams	98.8	98.3
A textbook chapter without diagrams	1.3	1.7

Values are the percentage of participants in each experiment endorsing the stated belief.

next one; no time limit was placed on the study of each paragraph, but participants were not able to go back to paragraphs that had already been read. After participants clicked this icon, the paragraph disappeared and then participants made a judgement from 0% to 100% for that paragraph, which indicated the likelihood that they would be able to correctly answer questions based on the information in the paragraph. All whole-number values from 0% to 100% were accepted, with "0%" indicating that a participant judged they definitely *would not* be able to answer questions on that paragraph correctly and "100%" indicating that a participant judged they definitely *would* be able to answer questions on that paragraph correctly. Intervening values indicated a corresponding likelihood of being able to answer the question correctly (e.g., 23% = "23% chance of getting the answer correct", 67% = "67% chance of getting the answer correct", etc.). After entering a judgement for a paragraph, the next paragraph was displayed. This study–judge pattern was repeated until all six paragraphs had been studied and judged. After the last paragraph had been judged, participants made a post-study global judgement concerning their overall understanding of how lightning storms develop on a 0 to 100 scale.

Participants were then tested on their learning of how lightning storms develop. They answered the 12 short-answer comprehension questions in a randomised order. They had an unlimited amount of time to answer each question but could not go back to a question once it had been answered (or left blank). After answering all of the questions, participants were debriefed.

Results

Beliefs questionnaire. The results of the BAML questionnaire (Table 1) indicate that the participants in Experiment 1 (as well as Experiment 2) strongly endorsed the belief that multimedia presentations (i.e., text with diagrams) tend to produce better learning than do single-medium presentations (i.e., text alone). This was true both for beliefs about the efficacy of multimedia in general and for the individual respondent.

Study time. The total number of seconds spent studying the materials (Table 2) was greater for the multimedia group than for the text-only

TABLE 2
Descriptive statistics for Experiment 1

	Text-only		*Multimedia*	
Measure	*M*	*(SD)*	*M*	*(SD)*
Study Time				
Total (seconds)	218.7	(81.3)	263.6	(112.5)
Test Performance				
% Correct	34.6	(22.7)	50.6	(24.8)
Judgements				
Pre-Study	63.8	(22.1)	70.5	(16.8)
Per-Paragraph	54.9	(20.2)	67.8	(19.4)
Post-Study	57.5	(20.0)	67.7	(18.6)

group, $F(1, 78) = 4.2$, $MSE = 9630.2$, $p = .04$, $\eta_p^2 = 0.05$. As described next, however, multimedia effects on learning obtained despite this difference in study time.

Test performance. Two raters each scored half of the participants' responses. Agreement was high (93%, *Kappa* = .85), so one rater scored all responses for analysis. Performance (Table 2) was greater for the multimedia group than for the text-only group, $F(1, 78) = 9.1$, $MSE = 564.9$, $p = .003$, $\eta_p^2 = 0.1$.

We also performed an ANCOVA using total study time as a covariate. Study time was significantly related to test performance, $F(1, 77) = 26.9$, $MSE = 423.9$, $p < .001$, $\eta_p^2 = 0.26$, but performance was still greater for the multimedia group than for the text-only group after controlling for the effect of study time, $F(1, 77) = 5.0$, $MSE = 423.9$, $p = .03$, $\eta_p^2 = 0.06$.

Judgement magnitude. The authors calculated the mean per-paragraph judgement for each participant across the six paragraphs, and then calculated the mean of these means across participants. These means and the mean global metacomprehension judgements (both pre- and post-study) for the two groups are presented in Table 2. Judgement magnitude was analysed with a 2 (format: text-only vs multimedia) × 3 (judgement type: pre-study global vs per-paragraph vs post-study global) mixed ANOVA. Judgement magnitude was greater for the multimedia group than for the text-only group, $F(1, 78) = 6.5$, $MSE = 905.8$, $p = .01$, $\eta_p^2 = 0.08$, and judgement magnitude differed by judgement type, $F(2, 156) = 6.2$, $MSE = 121.0$, $p = .002$, $\eta_p^2 = 0.07$, with the pre-study judgements being higher than either the per-paragraph judgements, $F(1, 78) = 9.1$, $MSE = 297.2$, $p = .003$, $\eta_p^2 = 0.1$, or the post-study judgements, $F(1, 78) = 5.0$, $MSE = 341.0$, $p = .03$, $\eta_p^2 = 0.06$. Most important, the two factors did not interact, $F(2, 156) = 1.6$, $MSE = 121.0$, $p = .2$, $\eta_p^2 = 0.02$, suggesting that the multimedia effect persisted across all judgements.[1]

Discussion

Almost all of the participants in Experiment 1 believed that multimedia presentations produce greater learning than does reading text alone (or even viewing diagrams alone), both for themselves and for people in general (Table 1). Although this outcome precluded the possibility that we could analyse participants within study conditions based on their *differing* beliefs about multimedia effect, the outcome is not surprising given that adding diagrams to text passages typically does improve learning (Mayer, 2005, 2009) and that students prefer multimedia materials to single-medium materials (Mayer & Massa, 2003).

This belief was consistently reflected in the judgements of text learning (Table 2), which suggests that participants used a multimedia heuristic. Using this heuristic, however, could potentially lead to errors if judgements were influenced more by these beliefs than by information about cognitive processes that directly influence performance (cf. McCabe & Castel, 2008; Nisbett & Wilson, 1977). In Experiment 1 this possibility could not be evaluated because a multimedia effect was found, which also poses problems for concluding that per-paragraph and post-study global judgements are based on a multimedia heuristic. In particular, both of these judgements may reflect on-line processing that boosts both judgements *and* test performance. This possibility can be explored only in situations where diagrams do not boost performance, because the consistent use of a multimedia heuristic should lead to judgement error, at least for the pre-study global judgements. However, if per-paragraph and post-study global judgements track performance and are not biased by a multimedia heuristic, then these judgements should not demonstrate a multimedia effect when diagrams do not boost performance. We evaluate this prediction in Experiment 2.

Finally, judgements decreased from the pre-study judgements to the judgements made during or after reading. This decrease may reflect participants' realisation that the task will be more difficult than they had expected (because how lightning develops is more complex than they had expected) or some other less interesting factor (e.g., fatigue or boredom). However, given that our focus is on the degree to which multimedia formats influence each kind of judgement,

[1] For interested readers, we calculated the relative accuracy of the per-paragraph judgements for the two groups (i.e., within-participants gamma correlations between judgements and test performance across paragraphs). Relative accuracy did not differ across groups in either Experiment 1 or 2 (Table B1 of Appendix B), so we do not discuss this measure further.

we will not discuss this general decrease in judgements any further and leave exploration of its causes for future investigations.

EXPERIMENT 2

In Experiment 2 we attempted to replicate the findings from Experiment 1, but most important, we added an additional group who received images that were not expected to boost test performance. These images were related to the topic (How lightning storms develop), but consisted of photographs of lightning strikes. To foreshadow, these photographs did not improve test performance, which allowed us to evaluate whether judgements were consistently higher for multimedia texts (regardless of whether or not multimedia increased test performance) or whether they would track performance.

Method

Participants, design, and materials. The participants were 120 undergraduates from Texas Tech University; they participated to partially fulfil a course requirement. The survey materials, study materials, and short-answer questions were the same as in Experiment 1 except that a new condition—text-with-photos—was added. This group (henceforth, the "photos group") received the same text as the other two groups, but each paragraph was accompanied by a different photo of lightning. An example photo is included in Figure A1 of Appendix A. The photos did not illustrate information in the text, but merely showed images of lightning strikes.

The format of the materials for study (text with diagrams, text alone, or text with photos) was the only between-participants independent variable. Participants were randomly assigned to one of the study conditions when they arrived at the laboratory, with the restriction that an equal number (40) were eventually assigned to each of the three groups. Regardless of their study condition, all participants later answered the same set of short-answer questions as a test of their understanding of the study materials as in Experiment 1.

Procedure. The procedure of Experiment 2 was basically the same as in Experiment 1. The only difference was the addition of the photos group that received a different photo of lightning with each paragraph (Figure A1).

Results

Beliefs questionnaire. Again, the results of the BAML questionnaire (Table 1) indicate that the participants in Experiment 2 strongly endorsed the belief that multimedia presentations tend to produce better learning than do single-medium presentations.

Study time. The total number of seconds spent studying the materials (Table 3) did not differ significantly between the three groups, $F(1, 117) = 2.2$, $MSE = 10445.9$, $p = .1$, $\eta_p^2 = 0.04$, although trends were evident (as in Experiment 1) indicating the participants spend nominally more time with the two multimedia formats than with the text-only format.

Test performance. Two raters each scored half of the participants' responses. Agreement was high (96%, $Kappa = .90$), so one rater scored all responses for analysis. Performance (Table 3) was different for the three groups, $F(2, 117) = 5.5$, $MSE = 524.0$, $p = .005$, $\eta_p^2 = 0.09$. As predicted, follow-up analyses (Tukey HSD) revealed that performance was highest for the standard multimedia group (as compared to the text-only group, $p = .005$, and to the photos group, $p = .052$). Performance did not differ for the text-only and photos groups ($p = .672$).

Although the photos did not add any information to the text paragraphs, this group did show a slightly elevated level of performance, which itself may reflect the nominally extra time they spent studying the texts (see Table 3). For this reason we also performed an ANCOVA using total study time as a covariate. Study time was significantly related to test performance, $F(1, 116) = 20.0$, $MSE = 451.0$, $p < .001$, $\eta_p^2 = 0.15$, but a significant effect of

TABLE 3
Descriptive statistics for Experiment 2

	Text-only		*Multimedia*		*Photos*	
Measure	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Study Time						
Total (seconds)	192.6	(89.7)	230.0	(97.4)	237.3	(117.5)
Test Performance						
% Correct	26.5	(19.8)	42.9	(26.4)	30.8	(22.0)
Judgements						
Pre-Study	65.3	(19.0)	74.1	(15.1)	75.1	(10.6)
Per-Paragraph	56.2	(18.3)	66.5	(17.7)	68.1	(17.5)
Post-Study	56.2	(21.7)	65.3	(20.9)	67.2	(18.0)

format still was evident after controlling for the effect of study time, $F(2, 116) = 4.9$, $MSE = 451.0$, $p = .009$, $\eta_p^2 = 0.08$. Contrasts revealed that, when study time was controlled for, test performance was greater for the standard multimedia group than for both the text-only ($p = .007$) and photos ($p = .009$) groups, and that the text-only and photos groups did not differ ($p = .914$). This produced the ideal conditions under which to compare the groups' judgements of text learning, because the photos themselves did not significantly affect learning.

Judgement magnitude. The means for per-paragraph judgements and the mean global judgements (both pre- and post-study) are presented in Table 3. Judgement magnitude was analysed with a 3 (format: text-only vs standard multimedia vs photos) × 3 (judgement type: pre-study global vs per-paragraph vs post-study global) mixed ANOVA. Judgement magnitude was different for the three groups, $F(2, 117) = 5.7$, $MSE = 735.7$, $p = .004$, $\eta_p^2 = 0.09$. Follow-up analyses (Tukey HSD) revealed that the text-only group tended to make lower judgements than did either the standard multimedia ($p = .022$) or photos group ($p = .006$). As important, the judgements of the standard multimedia and photos groups did not differ overall ($p = .903$). Judgement magnitude also differed by judgement type, $F(2, 234) = 24.2$, $MSE = 112.7$, $p < .001$, $\eta_p^2 = 0.2$, with the pre-study judgements being higher than either the per-paragraph judgements, $F(1, 117) = 29.0$, $MSE = 257.1$, $p < .001$, $\eta_p^2 = 0.2$, or the post-study judgements, $F(1, 117) = 27.1$, $MSE = 326.2$, $p < .001$, $\eta_p^2 = 0.2$. The two factors did not interact, $F(4, 234) = 0.1$, $MSE = 112.7$, $p = 1.0$, $\eta_p^2 = 0.002$, indicating that the influence of multimedia formats (either effective or ineffective ones) persists across the task.

Discussion

As in Experiment 1, comparing the judgements of the text-only and the two multimedia groups (standard and photos) in Experiment 2 provides evidence that people's beliefs about multimedia learning can affect metacomprehension judgements. The higher judgements for both multimedia presentations suggest that participants based their judgements in part—if not completely—on the belief that multimedia presentations tend to produce superior learning than does text alone (cf. Dunlosky & Hertzog, 2000). As previously noted, this conclusion is weakened by the fact that the standard multimedia group significantly outperformed the text-only group on the test, so it is not clear how much of the obtained difference stemmed from beliefs or from experience with the materials. Consider, however, that the photos group in Experiment 2 did not perform better than the text-only group, yet judgements from the photos groups were just as high as those made by the standard multimedia group. These outcomes provide stronger evidence that participants' judgements are based on a multimedia heuristic that emphasises the efficacy of multimedia materials more than the actual experience studying them.

GENERAL DISCUSSION

The effect of beliefs on metacomprehension judgements

A main goal of the present research was to evaluate whether people incorporate their beliefs about the effects of text format on learning into their judgements of text learning. To accomplish this goal we explored students' beliefs about multimedia materials and whether they influence metacomprehension judgements. Although the effects of multimedia materials on comprehension and learning per se have received a great deal of attention in the field (e.g., Mayer, 2005, 2009), metacomprehension within this domain has largely been unexplored. The two previous studies investigating multimedia and metacomprehension either reported non-significant trends in the expected direction (Waddill & McDaniel, 1992) or did not report judgement magnitude (Cuevas et al., 2002),[2] so their studies did not definitively

[2] Cuevas et al. (2002) examined between-participant correlations across participants, which tap a different aspect of judgement accuracy than do measures of within-participant relative accuracy that we reported in Table B1 (for a comparison of these measures, see Dunlosky & Hertzog, 2000). When we computed the between-participant correlations in our study as in Cuevas et al., they were statistically equivalent for multimedia materials versus for text alone. Thus the multimedia effects on this measure of accuracy reported by Cuevas et al. (2002) may not be highly robust or may be moderated by other factors, such as verbal ability (which we did not measure) or any other number of differences in methods between their experiment and the present ones. Given that this form of judgement accuracy was not vital to our present aims, we leave exploration of this inconsistency for future research.

address the question most central to the present research: Do beliefs about text materials influence judgements of text learning?

Based on the present evidence, the answer to this question is "yes". First, the results of the BAML questionnaire established that nearly all of the participants in our experiments believed that text with diagrams produces greater learning than does text alone. Second, in Experiment 1 participants' judgements of text learning were significantly greater when they were made for the multimedia version of the text than for the text-only version. These outcomes provided preliminary evidence that the participants' judgements reflected their belief that multimedia improves learning, especially for the pre-study judgements that could not be influenced by experience with the texts. One criticism of this conclusion pertains to the per-paragraph and post-study judgements—given that learning was better for the multimedia group in Experiment 1, these judgements might have reflected monitoring during study of how well the materials were learned, rather than the use of a multimedia heuristic.

Third, and most important, this possibility was disconfirmed by the results of Experiment 2. In that experiment the photos group did not outperform the text-only group, but the photos group judged their learning at all three times (pre-study, per-paragraph, and post-study) as being on the same level as that of the standard multimedia group. Thus, rather than being based on the online monitoring of comprehension, these judgements reflected people's beliefs about text materials—even when those beliefs were invalid.

As a whole, the outcomes of the present experiments are consistent with heuristic-based accounts of metacomprehension judgements in which people rely on heuristics to construct their assessments of learning text materials. In the present case participants over-relied on this heuristic, and this misapplication of a *multimedia heuristic* inappropriately biased their judgements in a situation—the photos group—where multimedia did not boost test performance (Glenberg et al., 1987). Of course, in many situations such a multimedia heuristic will be valid (Mayer, 2009), but its indiscriminate use will reduce the absolute accuracy of people's judgements of text learning in some situations. For this reason, one goal for future research will be to discover techniques that can align people's judgements more with learning levels.

As described earlier, one potential limitation with the present experiments concerns the effects of study time on test performance—participants tended to spend more time studying the multimedia formats than text alone (although such trends were not always significant), and this might have produced the differences in test performance rather than the actual format of the study materials. In contrast to this possibility, post-hoc analyses revealed that study time *was* predictive of learning, but also that any effects of study time were not responsible for the multimedia effects on test performance. Therefore the positive influence of multimedia is likely based more on processing differences between the two groups (Mayer, 2009) than on differences in study time. In fact, in some cases, study time with multimedia materials is *inversely* related to the instructional efficacy of the materials (e.g., Pollock, Chandler, & Sweller, 2002; Szabo & Kanuka, 1998), further demonstrating that it is more likely the content and format of the materials—not time spent studying them—that causes multimedia-learning effects.

Future directions

Wiley, Griffin, and Thiede (2005) warned researchers that subtle differences in text materials—such as whether a text is explanatory or merely provides a collection of ideas—can lead to inappropriately low estimates of people's metacomprehension accuracy. One conclusion that they offered was that "it seems problematic to run studies using different kinds of expository text if they are not all processed, represented, or understood in similar ways" (p. 420). We resonate with their arguments that research on metacomprehension should be informed by theories of text processing (cf. Rawson, Dunlosky, & Thiede, 2000) and should avoid confounds such as using a variety of text types within a single study. Nevertheless, it will also be critical to continue investigating how people evaluate their learning and comprehension of multiple kinds of expository text and text formats within a single reading session, partly because the kinds of lengthy text people read often include many kinds of exposition. For instance, a single chapter in an introductory psychology textbook will include explanations, collections of ideas, text with and without diagrams, photographs, charts, and so forth. Thus, we recommend that researchers are sensitive to these differences in text (cf. Wiley et al., 2005) and also directly assess how they can bias people's judgements.

Results from the present research indicate that promising directions for future research would be to develop questionnaires that tap people's beliefs about these formats and then systematically evaluate whether (and how) they relate to judgements of text learning (see also Hertzog et al., 2009). In the present experiments we administered the BAML questionnaire before participants read and judged their text learning. Although the reactive effects of strategy questionnaires are apparently small in some contexts (see e.g., Hertzog et al., 2009), asking participants about multimedia learning may have triggered beliefs and increased the chances that they would influence subsequent judgements. Evaluating this possibility and examining the factors that promote the use of belief-based heuristics on people's metacognitive judgements is an important area for future inquiry. Finally, if participants are asked to make judgements across phases of learning (e.g., before, during and after texts are read) as in the present research, researchers can systematically evaluate the joint influences of beliefs and on-line processing to people's judgements of text learning. When applied in the present context, people's beliefs about multimedia apparently dominated the construction of judgements at all time points (i.e., before, during, and after study), even when the multimedia format did not improve test performance.

Manuscript received 8 January 2009
Manuscript accepted 2 July 2010
First published online 19 August 2010

REFERENCES

Baker, J. M. C., & Dunlosky, J. (2006). Does momentary accessibility influence metacomprehension judgements? The influence of study–judgement lags on accessibility effects. *Psychonomic Bulletin & Review*, *13*, 60–65.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55–68.

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology, 34,* 30–41. doi: 10.1016/j.cedpsych.2008.07.001

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science, 30,* 433–464. doi: 10.1023/A:1020516301541

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15,* 462–474. doi: 10.1037//0882-7974.15,3.462

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16*, 228–232.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*, 906–911.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, *15*, 84–93.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*, 119–136.

Hacker, D. J. (1998). Self-regulated comprehension during normal reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 165–191). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Metamemory in adulthood: Differentiating knowledge, belief, and behavior. In T. M. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 161–212). Oxford, UK: North-Holland.

Hertzog, C., Dixon, R. A., Schulenberg, J. E., & Hultsch, D. F. (1987). On the differentiation of memory beliefs from memory knowledge: The factor structure of the Metamemory in Adulthood scale. *Experimental Aging Research*, *13*, 101–107.

Hertzog, C., Price, J., Burpee, B., Frentzel, J. F., Feldstein, S., & Dunlosky, J. (2009). Why do people show minimal knowledge updating with task experience: Inferential deficit or experimental artifact? *Quarterly Journal of Experimental Psychology, 62,* 155–173. doi: 10.1080/17470210701855520

Jee, B., Wiley, J., & Griffin, T. (2006). *Expertise and the illusion of comprehension*. Proceedings of the Annual Conference of the Cognitive Science Society.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgements of learning. *Journal of Experimental Psychology: General*, *126*, 349–370.

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition*, *34*, 959–972.

Lin, L. M., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, *23*, 345–391.

Maki, R. H. (1998). Text predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, *88*, 64–73.

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology, 95,* 833–841. doi: 10.1037/0022-0663.95.4.833

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, *43*, 31–43.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgements of scientific reasoning. *Cognition, 107,* 343–352. doi: 10.1016/j.cognition.2007.07.017

Metcalfe, J., & Finn, B. (2008). Evidence that judgements of learning are directly related to study choice. *Psychonomic Bulletin Review, 15,* 174–179. doi: 10.3758/PBR.15.1.174

Moore, D., Lin-Agler, L. M., & Zabrucky, K. M. (2005). A source of metacomprehension inaccuracy. *Reading Psychology*, *26*, 251–265.

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development*, *53*, 35–45.

Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgements. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 223–232.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12*, 61–86.

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 69–80. doi: 10.1037//0278-7393.28.1.69

Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The reading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, *28*, 1004–1010.

Rawson, K. A., & Kintsch, W. (2005). Rereading effects depend on time of test. *Journal of Educational Psychology 97,* 70–80. doi: 10.1037/0022-0663.97.1.70

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625.

Serra, M. J. (2010). Diagrams increase the recall of nondepicted text when understanding is also increased. *Psychonomic Bulletin & Review, 17,* 112–116. doi:10.3758/PBR.17.1.112

Serra, M. J., & Metcalfe, J. (2009). Effective implementation of metacognition. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 278–298). New York: Psychology Press.

Szabo, M., & Kanuka, H. (1998). Effects of violating screen design principles of balance, unity, and focus on recall learning, study time, and completion rate. *Journal of Educational Multimedia and Hypermedia*, *8*, 23–42.

Thiede, K. W. (1996). The relative importance of anticipated test format and anticipated test difficulty on performance. *Quarterly Journal of Experimental Psychology*, *49A*, 901–918.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology 95,* 66–73. doi: 10.1037/0022-0663.95.1.66

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85–106). New York: Psychology Press.

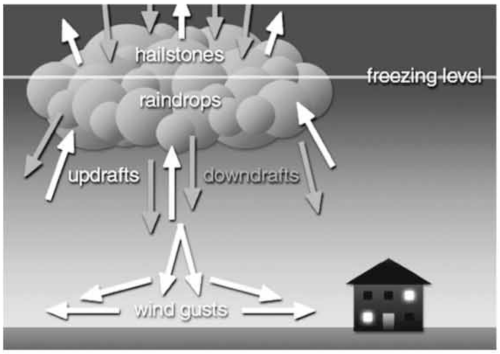
Waddill, P. J., & McDaniel, M. A. (1992). Pictorial enhancement of text memory: Limitations imposed by picture type and comprehension skill. *Memory & Cognition*, *20*, 472–482.

Weaver, C. A. III, Bryant, D. S., & Burns, K. D. (1995). Comprehension monitoring: Extensions of the Kintsch and van Dijk model. In: C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in Honor of Walter Kintsch* (pp. 177–193). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology*, *132*, 408–428.

Appendix A

Materials used in Experiments 1 and 2



Example Diagram

Example Photo

Example Paragraph

Eventually, the water droplets and ice crystals become too large to be suspended by updrafts. As raindrops and ice crystals fall through the cloud, they drag some of the air in the cloud downward, producing downdrafts. The rising and falling air currents within the cloud may cause hailstones to form. When downdrafts strike the ground, they spread out in all directions, producing gusts of cool wind people feel just before the start of the rain.

Example Questions

Q: Why do raindrops and ice crystals fall through the cloud?
A: They become too heavy to be suspended by updrafts.

Q: What produces gusts of cool wind people feel just before the start of the rain?
A: Downdrafts strike the ground and spread out in all directions.

Figure A1. Example paragraph, diagram, questions (Experiments 1 and 2), and photo (Experiment 2) used in the present experiments.

TABLE A1
Beliefs About Multimedia Learning (BAML) Questionnaire

Generalised Beliefs Questions

Which do you think is the most effective type of lesson in a science textbook for most people to learn from?

1. Just text
2. Mostly text with a few diagrams
3. About equal parts text and diagrams
4. Mostly diagrams with little text
5. A series of diagrams

Which do you think produces better learning in general, a science textbook chapter without any diagrams or a chapter with diagrams?

1. A textbook chapter without diagrams (i.e., just text)
2. A textbook chapter with diagrams (i.e., text with diagrams)

Belief About the Self Question

If you read a chapter in a science textbook that had diagrams to help you understand some sections but did not have diagrams for the other sections, which sections do you think you would understand better?

1. The sections without diagrams (i.e., just text)
2. The sections with diagrams (i.e., text with diagrams)

The response choices for each question appear after the question. Higher-numbered responses indicate more visually based responses.

Appendix B

Relative accuracy in Experiments 1 and 2

TABLE B1
Relative accuracy scores

	Text-only		*Multimedia*		*Photos*	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Experiment 1	+.23	(.62)	+.21	(.54)	–	–
Experiment 2	+.16	(.73)	+.12	(.58)	+.26	(.64)

Relative accuracy scores are the mean within-participant gamma correlation between per-paragraph metacomprehension judgements and test performance across paragraphs. Relative accuracy was not different across groups in either Experiment 1, $F(1, 67) = 0.03$, $MSE = 0.3$, $p = 0.9$, $\eta_p^2 = 0.0$, or in Experiment 2, $F(2, 96) = 0.4$, $MSE = 0.4$, $p = 0.7$, $\eta_p^2 = 0.0$.