Do Accurate Metacognitive Judgments Predict Successful Multimedia Learning?

Nicholas V. Mudrick (nvmudric@ncsu.edu)
North Carolina State University, Department of Psychology, 2310 Stinson Drive, 640 Poe Hall
Raleigh, NC 27695 USA

Michelle Taub (mtaub@ncsu.edu)
North Carolina State University, Department of Psychology, 2310 Stinson Drive, 640 Poe Hall
Raleigh, NC 27695 USA

Roger Azevedo (razeved@ncsu.edu)
North Carolina State University, Department of Psychology, 2310 Stinson Drive, 640 Poe Hall
Raleigh, NC 27695 USA

Abstract

Successful performance during multimedia learning requires accurate metacognitive judgments. However, little research has investigated the influence of accurate metacognitive judgments for different representations of information (e.g., text and diagram) on performance during multimedia learning. As such, we investigated if participants' metacognitive judgments for text and diagrams (i.e., content evaluations; CEs) were significantly related to increased performance and higher confidence during multimedia learning. Metacognitive judgments and performance measures were collected from 48 undergraduate participants during 18 randomized trials. Results using multilevel modeling indicated that participants' CEs for text-based content were significantly predictive of performance. Results also showed that accurate CEs for diagrams interacted with accurate multiple-choice responses to predict higher retrospective confidence judgments (i.e., higher confidence). Identifying metacognitive judgments predictive of increased performance during multimedia learning has important theoretical, conceptual, and analytical implications.

Keywords: multimedia learning; metacognition; meta-cognitive judgments; multilevel modeling; performance; science learning

Research indicates learning with multimedia materials (e.g., text and diagram) is more effective than learning through text alone (Butcher, 2014; Mayer, 2014). Successful multimedia learning entails individuals actively and accurately selecting, organizing, and integrating text- and image-based information into a coherent mental model (Mayer, 2014). However, research suggests learners do not always engage in accurate and effective metacognitive monitoring and regulation during learning with multimedia (Azevedo, 2014). Specifically, research has indicated participants often exhibit overconfidence when monitoring their own understanding during multimedia learning (Serra & Dunlosky, 2010). The multimedia heuristic suggests learners' own judgments of learning (JOLs; i.e., how well they will remember the information) are largely inflated when compared to their actual performance because individuals perceive multimedia content as being easier to learn than with text alone (Serra & Dunslosky, 2010).

Research on metacognitive monitoring during multimedia learning has traditionally employed modified meta-comprehension paradigms (based on Nelson & Narens' metamemory framework, 1990), during which participants are asked to make metacognitive judgments (e.g., ease-of-learning [EOL], immediate and delayed JOLs, retrospective confidence judgments [RCJs]) during various stages of multimedia learning (e.g., Burkett & Azevedo, 2012; Eitel, 2016; Pilegard & Mayer, 2015). The major assumption of this research is that the timing of metacognitive judgments made during multimedia learning (before learning, during learning, and after learning) will vary in accuracy, selection of cognitive strategies, and subsequent performance, dependent on the specific experimental manipulation (e.g., delayed JOLs are more predictive of performance than EOLs; Burkett & Azevedo, 2012; Nelson & Dunlosky, 1991). As this research has identified that most meta-cognitive judgments for multimedia are often inaccurate (e.g., Serra & Dunlosky, 2010), much of the literature has focused on ways to improve metacognitive judgments. For example, some research has focused on manipulating the framing of metacognitive judgment prompts to improve judgment accuracy (e.g., Pilegard & Mayer, 2015; Vössing, Stamov-Roßnagel, & Heinitz, 2016). Pilegard and Mayer (2015) compared JOLs (i.e., *how well do you remember the content*) to judgments of understanding (JOUs; i.e., *how well do you understand the information*) and found JOUs were more predictive of retention and transfer compared to JOLs. These findings suggest that framing metacognitive judgment prompts (e.g., from JOLs to JOUs) significantly impacts the metacognitive processes employed during multimedia learning, potentially indicating there may be other metacognitive judgments participants use that can successfully influence performance. In support of this assertion, research on hypermedia and self-regulated learning (SRL) suggests several other metacognitive processes may be more predictive of multimedia learning outcomes (Greene & Azevedo, 2009).

Azevedo, Greene, and Moos (2007) developed a classification scheme by which 35 micro-level meta-cognitive judgments can be evident during successful SRL with hypermedia-based learning environments. One

example of these judgments is a content evaluation (CE). CEs are judgments learners make to assess the relevancy of the content (e.g., multimedia) they are viewing to their current goal (e.g., answering a science question about a human body system; Greene & Azevedo, 2009). CEs are key metacognitive judgments for successful multimedia learning, such that accurate CEs can direct participants to study more efficiently. For example, if the goal is to answer a science question about the human body system and participants evaluate the text but not the diagram they are viewing to be relevant to their goal, they should invest more effort and time to study the text (as opposed to the diagram), employ the appropriate cognitive strategy (e.g., make an inference), and therefore be more likely to answer the question correctly.

Other research on metacognitive judgments during hypermedia learning has identified the predictive validity of traditional metacomprehension judgments like RCJs. For example, Mengelkamp and Bannert (2010) investigated the stability of participants' RCJs as they learned about operant conditioning with a hypermedia environment. Results indicated that the absolute accuracy (i.e., difference between judgments and performance) was stable throughout the learning session, and relative accuracy (correlation between judgments and performance) was significantly predictive of hypermedia learning outcomes.

Theories of multimedia learning suggest participants cognitively process information from text and diagrams separately and in different ways (Burkett & Azevedo, 2012; Mayer, 2014). Additionally, researchers have outlined the multimedia effect to indicate that students demonstrate longer periods of recall and higher levels of retention when learning with text and images as opposed to learning only with text (Butcher, 2014). However, evidence suggests learners do not always engage in effective selection, organization, and integration of multiple representations and instead exhibit a bias toward text-based (as opposed to diagram-based) information during multimedia learning (Hegarty & Just, 1993). Since cognitive processes are different for text and diagrams, it should be expected that metacognitive judgments will also be different.

Accurate metacognitive monitoring and regulation are required during multimedia learning to achieve an increase in learning outcomes (Azevedo, 2014). However, little research has examined the specific processes underlying successful metacognitive monitoring and regulation during multimedia learning. Specifically, few metacognitive judgments have been found to be predictive of successful multimedia learning outcomes (e.g., overconfident JOLs; Serra & Dunlosky, 2010). We argue that examining other metacognitive judgments (CEs, RCJs) can inform us of monitoring processes that are more indicative of successful learning and performance. In contrast to the limited research on metacognitive judgments during multimedia learning, we focus on different metacognitive judgments and identify how they can contribute to superior learning outcomes.

In this study, we examined participants' text CEs, diagram CEs, multiple-choice responses, and RCJs during multimedia learning to answer the following three questions: (1) Are accurate text and diagram CEs associated with an increase in the likelihood of an accurate multiple-choice response? (2) Is there a significant relationship between text and diagram CE accuracy and RCJs? (3) Is there a significant relationship between the interactions of text and diagram CEs and multiple-choice responses on RCJs?

To address our research questions, we proposed the following hypotheses:

H1: Accurate text and diagram CEs will be significantly associated with an increase in the likelihood of an accurate multiple-choice response.

H2: The relationship between text and diagram CE accuracy and RCJs will be significant.

H3: The relationship between the interactions of text and diagram CEs and multiple-choice responses on RCJs will be significant.

Method

Participants

Forty-eight undergraduates (69% female) enrolled at a large mid-Atlantic university participated in this study. Their ages ranged from 18 to 24 ($M = 20.04$, $SD = 1.60$), and they were compensated up to $30 for their participation.

Experimental Design

This study used a 3×3×2 within-subjects design (18 trials). Each participant was exposed to three *human agent facial expressions*: neutral (neutral facial expression), congruent (i.e., joy for facial expressions congruent with the content relevancy), and incongruent (i.e., confusion for facial expressions incongruent with content relevancy). Each participant was also exposed to each type of multimedia *content relevancy*: fully relevant (text and diagram relevant to the question), text somewhat relevant (but diagram still fully relevant), and diagram somewhat relevant (but text still fully relevant). Additionally, *two types of questions* were posed: function (regarding the function of a body system) and malfunction (regarding a malfunction of a body system). Based on these manipulations each student completed 18 trials, with different combinations of human agent facial expression, multimedia relevancy type, and question type. For this paper, our analyses focused on meta-cognitive judgments across the trials and experimental manipulations.

Materials

The materials used in this study included the following: an informed consent form; a demographic questionnaire; and a researcher-developed, 4-foil, 18-item multiple-choice pretest of basic knowledge of human body systems (e.g., integumentary and nervous systems). Each question on the

pretest specifically related to the content presented in each multimedia science content slide.

Additionally, this study included 18 researcher-developed multimedia science content slides developed with a faculty member in human biology. The relevancy manipulations were created by including information that was related to but not necessary for answering the question.

MetaTutor Multimedia Learning Environment

The MetaTutor multimedia learning environment is a multimedia-based content presentation tool designed to examine the influence of a human agent's facial expressions on participants' cognitive strategies and metacognitive judgments during learning about human body systems. The environment consists of a human agent capable of facially expressing several emotional states (i.e., neutral, confusion, joy), science questions and corresponding multimedia science content, and metacognitive judgment prompts (EOLs, text and diagram CEs, and RCJs). The multimedia science content consists of three paragraphs (Flesch-Kincaid readability score range: 9.1–12.5; $M = 10.5$) and a diagram depicting the concept described in the text.

The environment presents 18 linearly structured, self-paced trials consisting of metacognitive judgments (e.g., EOLs, CEs, and RCJs), multimedia content presentation, and human agent facial expressions.

The 18 trials have the identical format. In each trial, participants are first presented with a science question and asked to submit an EOL, *How easy do you think it will be to learn the information needed to answer this question?* Participants made their EOL judgment on a scale from 0% to 100%, increasing in increments of 20%. Participants were then presented with a content slide containing the text, diagram, science question presented previously, and human agent. After 30 s (to ensure participants had enough time to initially review the material), participants were prompted to assess the relevancy of both the text and diagram, *Do you feel the text/diagram on this page is relevant to the question being asked?*, by making two CE judgments on a Likert-type scale (ranging from 1–3) on the following statements: *The text/diagram is relevant*, *The text/diagram is somewhat relevant*, and *The text/diagram is not relevant*. Upon making their text and diagram CEs, the human agent expressed a congruent, incongruent, or neutral facial expression based on the relevancy of the content (e.g., a congruent facial expression of joy if the text and diagram were relevant to the question being asked). Following the agent's expression, participants were permitted to reread the text and reinspect the diagram at their own pace. After they re-examined the multimedia content, participants were prompted to answer the science question by choosing the correct response from 4-foil answers. After submitting their answer, participants were prompted to make a RCJ by answering *How confident are you that the answer you provided is correct?* Participants made their judgment on a scale from 50% to 100% increasing in increments of 10%. After submitting their response, participants were required to justify their

answer by typing their response into a text box. Subsequently, participants were asked to make another RCJ based on their justification. This procedure was followed for all 18 trials with each trial randomized across participants.

Procedure

Once participants entered the lab they were asked to complete an informed consent form. Then the eye tracker was calibrated by the researcher.[1] Following calibration, participants were asked to complete a computerized demographic questionnaire and an 18-question, 4-foil pretest that assessed their basic science knowledge across the multiple body systems (e.g., urinary, endocrine) presented in the experiment. After the pretest, participants completed the 18 previously described trials. The experimental session lasted approximately 90 min.

Coding

Text and diagram CE judgments were recorded across the 18 trials (i.e., 18 text + 18 diagram = 36 total CE judgments for each participant). Responses were coded based on their accuracy, such that an accurate CE judgment was given a score of 1, a partially correct judgment was scored as 0.5, and an incorrect judgment was scored as 0. For example, if participants judged the diagram as somewhat relevant and a text as fully relevant during a "diagram somewhat relevant" trial, they were given a score of 1 for each response because the text was still fully relevant to the question being asked, whereas the diagram was only somewhat relevant.

Participants' responses to the 4-foil, multiple-choice questions were coded by correctness. A correct response was coded as 1 and an incorrect response was coded as 0.

Participants' RCJs were coded on a scale from 50% to 100%. A score of 50% indicated participants simply guessed at their answer (indicating they believed they had a 50/50 chance of getting their answer correct), whereas a score of 100% indicated participants were completely confident in their response.

Results

Research Question 1: Are accurate text and diagram CEs associated with an increase in the likelihood of an accurate multiple-choice response?

A fully unconditional model (i.e., with no predictor variables) dichotomous outcomes (i.e., accurate multiple-choice response = 1, inaccurate = 0), was conducted on multiple-choice accuracy. Results indicated that the average probability of responding to a multiple-choice question correctly was 60%.

A dichotomous outcomes model was conducted on multiple-choice accuracy (i.e., accurate = 1, inaccurate = 0) with text and diagram CE accuracy as the predictor variables. Results revealed that more accurate text CEs

---

[1] Although eye-tracking data were collected, they were not analyzed for this study.

($OR = 1.98$, $t = 3.09$, $p = 0.002$) but not diagram CEs ($OR = 0.98$, $t = -0.10$, $p > 0.5$) were associated with an increase in the likelihood of correctly answering multiple-choice questions. Specifically, as text CE response accuracy increased, there was a 98% increased chance of responding correctly. That is, if participants were accurate in their text CEs, they were substantially more likely to respond correctly to the multiple-choice questions.

Research Question 2: Is there a significant relationship between text and diagram CE accuracy and RCJs?

A fully unconditional model conducted on RCJs indicated 29.8% of the variability was between participants ($\tau_{00} = 79.61$, $z = 4.24$, $p < 0.001$) and 70.2% was within participants ($\sigma^2 = 187.49$, $z = 19.99$, $p < 0.001$), justifying further analysis.

An unconstrained multiple level 1 predictor model was run on RCJs using text CE and diagram CE accuracies as the predictor variables. Results revealed that an increase in both text CE accuracy ($\gamma_{10} = 5.70$, $t = 3.95$, $p < 0.001$) and diagram CE accuracy ($\gamma_{20} = 6.01$, $t = 4.63$, $p < 0.001$) significantly predicted an increase in RCJs. As the accuracies of participants' text and diagram CEs increased, their reported confidence in their performance also increased. This model accounted for 6.2% of the within-participant variance in participants' RCJs.

Research Question 3: Is there a significant relationship between the interactions of text and diagram CEs and multiple-choice responses on RCJs?

A constrained multiple level 1 predictor model was run on RCJs using text and diagram CE accuracies and their interactions with multiple-choice responses as predictor variables. Results indicated the interaction between text CE accuracy and multiple-choice response accuracy was not significant ($\gamma_{40} = 1.50$, $t = 0.50$, $p = 0.62$). However, results did reveal a significant interaction effect between diagram CE accuracy and multiple-choice response ($\gamma_{40} = -7.21$, $t = -2.75$, $p = 0.006$), such that participants whose diagram CEs were most accurate and who also had more accurate multiple-choice responses also reported more confidence in their answers (see Figure 1). This model accounted for 7.7% of the within-participant variance in participants' RCJs.

Discussion

The goal of this study was to examine the relationships between metacognitive judgments and their contributions to increased performance during multimedia learning. Overall, results revealed that when participants made accurate text CEs, they were more likely to respond correctly to multiple-choice questions. Additionally, accurate text and diagram CEs contributed to higher reported confidence in answers. As such, our findings augment current understanding of how different metacognitive judgments, from those
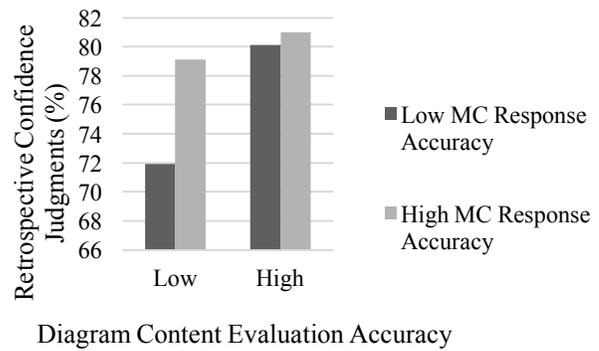

Figure 1: Interaction between diagram CE response accuracy and MC response accuracy on RCJs.

traditionally examined in the multimedia learning literature (e.g., JOLs), can contribute to improved performance and higher confidence.

Results from Research Question 1 indicated accurate text CEs were significantly predictive of an increased chance of responding correctly to multiple-choice questions, whereas diagram CEs were not. These results partially support our hypothesis, demonstrating participants could more accurately assess the relevancy of the text-based (as opposed to diagram-based) material related to answering the science question. Furthermore, these results are consistent with theories of multimedia learning that suggest individuals cognitively process text- and diagram-based material separately (Mayer, 2014; Schnotz, 2014). It is possible that participants not only cognitively process the text and diagrams separately, but also metacognitively monitor the information in text and diagrams separately and with varying levels of accuracy. Given evidence suggesting individuals exhibit a bias toward processing text-based information (at the expense of diagrams; Hegarty & Just, 1993), in addition to the redundancy of the diagram-based information to the text, participants may have realized the text-based information was sufficient and thus relevant enough to answer the multiple-choice questions correctly.

As hypothesized, results from Research Question 2 demonstrated that text and diagram CEs significantly predicted higher RCJs. Specifically, the more accurate participants' text and diagram CEs were, the more confident they were in their multiple-choice responses. Taken together with the previous finding, these results indicate participants may have relied on their relevancy judgments of both the text and diagram when they made their RCJs (as opposed to answering the question). As such, this finding significantly augments research on metacognitive judgments during multimedia learning by indicating a significant relationship between multiple metacognitive judgments.

Lastly, results from Research Question 3 indicated the interaction between diagram CE accuracy and multiple-choice response accuracy significantly predicted increased RCJs. More specifically, participants who provided more

accurate diagram CEs and responded accurately to multiple-choice questions also reported more confidence in their answers. These results partially support our hypothesis that both text and diagram CEs interact with multiple-choice responses to predict increased RCJs. Additionally, this result is supported by previous literature that suggests a significant relationship between performance and RCJs (e.g., Mengelkamp & Bannert, 2010). These results also support our assumption that since cognitive processes are different for different representations of information, so too are metacognitive monitoring processes. However, research is limited regarding the metacognitive processes involved when learning with and comprehending diagrams.

Overall, these results suggest that accurately assessing the relevancies of text and diagrams differentially impacts performance and future metacognitive judgments (e.g., accurate CEs related to increased RCJs). Results also indicated that when participants responded to multiple-choice questions, they relied on their metacognitive judgments of the text rather than diagrams. In contrast, participants relied on metacognitive judgments of diagrams and their performance when making RCJs. Previous research has indicated a significant relationship between CEs and performance (e.g., Greene & Azevedo, 2009). However, unlike previous literature, these results suggest text and diagram CEs differentially impact not only performance, but also reported confidence. Ultimately, these results confirm that other metacognitive judgments for different representations of information can predict greater performance during multimedia learning.

Limitations

Our study has several limitations. First, as we were primarily interested in the relationship between metacognitive judgments (e.g., CEs, RCJs) and performance across conditions, we did not examine the impact of content relevancy (e.g., fully relevant text and diagram, text less relevant, diagram less relevant) or question type (e.g., function vs. malfunction science question). Furthermore, the information needed to answer the multiple-choice questions correctly was primarily located in the text, which may have influenced participants' CE judgments. Future research should include separate function and malfunction questions based on the information presented in the diagrams. Moreover, we did not examine the accuracies of RCJs as multiple-choice responses were dichotomously coded as correct or incorrect. Future research will include measures of absolute and relative accuracies for RCJs (e.g., Schraw, 2009). Lastly, we can only make limited conclusions regarding the underlying cognitive and metacognitive processes (e.g., multiple fixations on irrelevant diagrams) that contributed to the accuracies of the text and diagram CEs and multiple-choice responses, as multichannel trace data (e.g., eye tracking) were not analyzed. Despite these limitations, this study has several important implications.

Future Directions and Implications

The results of this study have important implications for future studies examining the influence of metacognitive judgments on performance during multimedia learning. First, future research should include analyses of multi-channel trace data (e.g., eye tracking, facial expressions of emotions) that would allow for a more comprehensive depiction of the cognitive, affective, and metacognitive processes that occur when making CEs during multimedia learning (see Azevedo, 2014). Specifically, analyzing eye-tracking data can provide a micro-level description of the cognitive processes (e.g., coordination of information sources) contributing to increased performance and accurate text and diagram CEs. For example, does more time spent reading the text contribute to more accurate text CEs? Do specific eye-movement "signatures," as evidenced by scan path analyses, indicate greater integration of multimedia information and subsequently lead to increased performance? Further, examining the influence of participants' affective processes (e.g., emotions) would provide evidence of how they influence cognitive and metacognitive processes. For example, are participants' facial expressions of confusion predictive of decreased CE accuracy? How do participants' facial expressions of frustration influence the quality of their multiple-choice responses? Lastly, as this study was limited to analyzing the accuracy of RCJs, future research should seek to determine how CEs contribute to the accuracy of RCJs. It is possible that participants' CEs were accurate, but they exhibited over- or under-confidence when making their RCJs.

As our results indicated that diagram but not text CEs interacted with multiple-choice responses to predict RCJs, they emphasize the differential impact of multiple representations of information on participants' meta-cognitive judgments. Future research should examine the specific impact of different representations (e.g., diagrams, graphs, illustrations) on participants' metacognitive judgments to address the gap in the literature and gain better understanding of the metacognitive monitoring processes involved during multimedia learning.

Using a within-subjects design allowed us to examine the differential impact of how accurate metacognitive judgments influenced performance and confidence with reduced error caused by individual differences. Additionally, using multilevel modeling (Raudenbush & Bryk, 2002) enabled us to accurately assess within-subjects variance without violating traditional statistical assumptions (e.g., independence of observations) that many within-subjects designs ignore. Despite these benefits, future research should explore other experimental designs that are less controlled (e.g., more naturalistic) to increase the ecological validity of these findings. Due to our sample size, we did not find significant between-subjects variance; future research should replicate these analyses with larger samples to determine individual differences indicative of improved metacognitive judgment accuracy and performance (e.g., prior knowledge of body systems).

Additionally, these results indicate the importance of coordinating multiple sources of information (e.g., text and diagram) and can be used to inform the design of educational training regimens. For example, future research should explore the impact of cognitive (e.g., Bergey, Cromley, & Newcombe, 2015) and metacognitive (e.g., Azevedo, 2014) instruction that emphasizes how individuals should learn using both text and diagrams. Training can be provided to demonstrate how to accurately judge the relevancy of texts and diagrams, as well as emphasize the importance of accurate metacognitive judgments in relation to increased performance. Furthermore, these results can also inform the design of future intelligent, adaptive multimedia-based learning environments to support and scaffold accurate metacognitive judgments. If participants continuously make inaccurate text CEs, the system can intervene by cueing their attention to the relevant text-based information or by providing additional relevant declarative and conditional knowledge (e.g., how to accurately judge the relevancy of different representations of information).

Lastly, the results from this study suggest accurate metacognitive judgments are required for increased performance and confidence during multimedia learning. Traditionally, metacognitive judgments during multimedia learning have been found to be largely inaccurate. However, our results indicate other metacognitive processes (e.g., CEs) may be more informative of increased performance. For example, future studies could examine the influence of accurate feelings of knowing (i.e., individuals are aware of having read information but are unable to recall it on demand) and how they can contribute to increased performance during multimedia learning. As such, future research examining the influence of other metacognitive judgments will significantly augment our understanding—as well as the contemporary theoretical frameworks of multimedia learning—of the relationship between cognitive and metacognitive processes contributing to increased performance.

Acknowledgments

References

Azevedo, R. (2014). Multimedia learning of metacognitive strategies. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Azevedo, R., Greene, J., & Moos, D. (2007). The effect of a human agent's external regulation upon college students' hypermedia learning. *Metacognition and Learning*, *2*, 67–87.

Bergey, B. W., Cromley, J. G., & Newcombe, N. S. (2015). Teaching high school biology students to coordinate text and diagrams: Relations with transfer, effort, and spatial skill. *International Journal of Science Education*, *37*, 2476–2502.

Burkett, C., & Azevedo, R. (2012). The effect of multimedia discrepancies on metacognitive judgments. *Computers in Human Behavior*, *28*, 1276–1285.

Butcher, K. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Eitel, A. (2016). How repeated studying and testing affects multimedia learning: Evidence for adaption to task demands. *Learning and Instruction*, *41*, 70–84.

Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Education Psychology*, *34*, 18–29.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32,* 717-742.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge, England: Cambridge University Press.

Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: Stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition*, *38*, 441–451.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, *2*, 267–270.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26,* 125-173.

Pilegard, C., & Mayer, R. E. (2015). Within-subject and between-subject conceptions of metacomprehension accuracy. *Learning and Individual Differences*, *41*, 54–61.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.

Schnotz, W. (2014). The integrated model of text and graphics comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*, 33–45.

Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgments reflect the belief diagrams improve learning from text. *Memory*, *18*, 698–711.

Vössing, J., Stamov-Roßnagel, C., & Heinitz, K. (2016). Images in computer-supported learning: Increasing their benefits for metacomprehension through judgments of learning. *Computers in Human Behavior*, *57*, 221–230.