Can a Composite Metacognitive Judgment Accuracy Score Successfully Capture Performance Variance during Multimedia Learning?

Megan D. Wiedbusch[1] (meganwiedbusch@knights.ucf.edu)
Roger Azevedo[1] (roger.azevedo@ucf.edu)
Micheal J. Brown[2] (brown169@knights.ucf.edu)

[1]Department of Learning Sciences and Educational Research, 4364 Scorpius St, Orlando, FL 32816 USA
[2]Department of Computer Science, 4328 Scorpius St, Orlando, FL 32816 USA

Abstract

Theoretical models of self-regulated learning highlight the importance and dynamic nature of metacognitive monitoring and regulation. However, traditional research typically has not examined how different judgments, or the relative timing of those judgments, influence each other, especially in complex learning environments. We compared six statistical models of performance of undergraduates (n = 55) learning in MetaTutor-IVH, a multimedia learning environment. Three types of prompted metacognitive judgments (ease of learning [EOL] judgments, content evaluations [CEs], and retrospective confidence judgments [RCJs]) were used as individual predictors, and combined in a uniformly-weighted composite score and empirically based weighted composite score across the learning session. The uniformly weighted composite score better captured performance than the models using only an EOL judgment or RCJ judgment. However, the empirically weighted composite model outperformed all other models. Our results suggest that metacognitive judgments should not be considered as independent phenomenon but as an intricate and interconnected process.
Keywords: Learning; Metacognition; Self-regulated learning; Intelligent Tutoring Systems; Multimedia

Introduction

Self-regulated learning (SRL) involves learners actively monitoring, assessing, and modulating cognitive, affective, metacognitive, and motivational processes to accomplish learning objectives (Azevedo, Taub, & Mudrick, 2018; Dunlosky & Rawson, 2019). Research has consistently shown that effectively employing SRL processes (e.g., content evaluations) and strategies (e.g., note taking), improves academic performance, particularly when learning about complex topics and problem-solving tasks (e.g., Azevedo, 2014; Azevedo & Cromley, 2004; Bannert, Hildebrand, & Mengelkamp, 2009; de Boer, Kostons, & M., 2012; Dignath & Büttner, 2008; Jansen, Leeuwen, Janssen, & Jak, 2019). Many theoretical models of SRL highlight the importance and temporally dynamic nature of metacognitive monitoring and regulation (Schunk & Greene, 2017; Usher & Schunk, 2018; Winne, 2018).

Theoretical Background

Traditional research has examined metacognitive monitoring judgments using paradigms based on Nelson and Narens (1990)'s metamemory framework. According to this framework, learners initiate a range of metacognitive judgments at various temporal and conceptual stages of learning (Winne, 2018). While there is empirical evidence that earlier judgments (e.g., an EOL on the first trial) predict future performance of the same type of judgment (e.g., an EOL on the next trial; Tauber & Rhodes, 2012; Serra & Ariel, 2014), only recently have researchers begun exploring relations between different types of judgments. Chua and Solinger (2015) found that feeling-of-knowing judgments influenced RCJs. Additionally, Dougherty, Robey, and Buttaccio (2018) found that the inclusion of a judgment of learning could improve a subsequent RCJs. More research is still needed to understand the relationship of all metacognitive judgments. There is a need to understand how they impact learning and performance together. As such, our study aims to address this gap by comparing three types of judgments, (1) ease of learning (EOL), (2) content evaluations (CE), and (3) retrospective confidence judgements (RCJ)s, as independent explanatory factors and in conjunction as a composite of metacognitive monitoring to assess the extent to which these processes best explains performance (i.e., separately or together).

Ease of learning (EOL) judgments, or the initial assessment of how easy something will be to learn made in advanced of instructed study, are thought to be made through inferences from prior context and domain knowledge (Jemstedt, Kubik, & Jönsson, 2017). They are assumed to guide how we study by making decision on how much effort and attention to allocate, but have been found to be poor or sometimes only moderate predictors of material difficulty. This has been attributed to the lack of context missing prior to learning (Leonesio & Nelson, 1990; Mazzoni, Cornoldi, Tomat, & Vecchi, 1997; McCarley & Gosney, 2005; Son & Metcalfe, 2000; Britton, Van Dusen, Gülgöz, Glynn, & Sharp, 1991; Jönsson & Kerimi, 2011). However, Jemstedt et al. (2017) found EOLs could be accurate given high item variation, grading with a binary criterion, task type, and item presentation timing, especially within complex learning environments where we cannot assume a learner has zero prior knowledge.

Content evaluations (CEs) are a judgment measured during the learning process. First introduced in Greene and Azevedo (2007)'s SRL framework, an adaption of Winne and Hadwin (1998, 2012)'s and Pintrich et al. (2000)'s frameworks, they are the monitoring of content relative to goals. For example, a learner might read an introductory paragraph and conclude it is not relevant towards their learning goal and therefore decide to skip that section of the reading. Accurate CEs have been positively related to knowledge acquisition (Pilegard & Mayer, 2015) and the eye-gaze behaviors of learners (Dever, Wiedbusch, & Azevedo, 2019).

Occurring after learning, retrospective confidence judgments (RCJs) are the reported likelihood or confidence that a learner accurately recalled the tested information(Stretch & Wixted, 1998; Brewer & Sampaio, 2012; Ranganath et al., 2004; Chua, Schacter, Rand-Giovannetti, & Sperling, 2006). RCJs have been shown to be better predictors of learning compared to judgements that occur during the learning session (i.e., EOLs; (Dougherty, Scheck, Nelson, & Narens, 2005; Hines, Touron, & Hertzog, 2009; Ryals, Rogers, Gross, Polnaszek, & Voss, 2016)).

Current Study

The aim of our study was to explore how various metacognitive judgement accuracy could be captured with multiple self-reports temporally spaced to explain performance variance of a multimedia learning task. Previous research has shown that more accurate metacognitive judgments can significantly impact problem solving and learning about complex topics (Azevedo, 2014; Mayer, 2014; Taub & Azevedo, 2018), however most rely on a singular judgment type measured at a single point during the learning session. This could fail to capture the temporal fluctuations of metacognition that occur and the potential compounding interactions between judgments. We address this issue by posing the following research questions:

(1) Which metacognitive judgment best captures performance variation during learning with MetaTutor-IVH? Based on previous research (Leonesio & Nelson, 1990), we predict that EOLs will perform the worst as an explanatory variable. Furthermore, we believe that RCJs taken after reflection will outperform RCJs prior to reflection. Because CEs occur during the learning session, we predict that they will perform the best as they are not susceptible to retrieval effects like RCJs, but also occur once information and context is provided (unlike EOLs). This would follow similar patterns seen between judgments of learning (which also occurring during learning) and RCJs (Dougherty et al., 2018).

(2) How does a statistical model using a composite score compare to models only using individual components? We predict the composite score will outperform any model with only a single judgment predictor because of the assumption that regulation is cyclical and adaptive in nature. By capturing more of the learning session across the temporal scale, we believe the composite score will better reflect learning.

(3) How does an equally weighted composite score statistical model compare to a model weighted by our findings in (RQ2)? We predict that the score accounting for the strength of the explanatory relationship will outperform the equally weighted statistical model as it accounts for differences in judgment types.

Methods

Participants and Materials

Fifty-five undergraduates, 65% female, from a large North American University participated in our study. Their ages ranged from 18 to 30 (M = 20.38, SD = 2.58). All participants completed a demographic questionnaire, an 18-item human biological systems content multiple choice pretest, and several questionnaires assessing emotions and motivation (i.e., Achievement Emotions Questionnaire (AEQ; (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011)), Emotion Regulation Questionnaire (ERQ; (Gross & John, 2003)), Perceived Affect Utility Scale (PAUSe; (Chow & Berenbaum, 2012)). The main experimental study occurred in the MetaTutor-IVH environment. All participants were compensated $10/hr up to $30 for their participation. IRB approval was received prior to recruitment and data collection.

MetaTutor-IVH

The MetaTutor-IVH environment is a linearly structured multimedia learning environment designed to study prompted metacognitive judgments while learning about 9 human biology systems (Azevedo, Mudrick, Taub, & Bradbury, 2019). Learners go through 18 trials that follow an identical linear format at their own pace (see Figure 1). They are first presented with a science question and asked to submit an EOL (i.e., *"How easy do you think it will be to learn the information needed answer this question?"*) on a 1-unit sliding scale from 0 to 100. Following this judgment, they are presented with a multimedia content slide (see Figure 2) which consists of 3 paragraphs (Flesch-Kincaid readability score range: 9.1-12.5; M = 10.5) along the left side panel, a diagram in the center, the science question along the top, and an artificial agent in the top right hand corner. The content provided on the screen was developed with a biology expert and designed to sometimes contain not fully relevant information for the posed question. After 30 seconds on the content slide, the environment prompted CEs. Participants provided two CEs about the relevancy of (1) the text and (2) the diagram presented within each trial. They answered the question *"Do you feel the text/diagram on this page is relevant to the question being asked?"* on a 3-point rating scale (i.e., text/diagram is relevant; text/diagram is somewhat relevant; or text/diagram is not relevant). These judgments were then compared to the experimentally manipulated relevance of the content. Participants could choose when to respond to this prompt at any time. Once they submitted their judgment, for the next 10 seconds, the agent expressed its own judgment about the content relevancy (i.e., if it felt the content was relevant). Participants could continue studying the content slide until they were ready to answer the posed question. Once participants answered the multiple

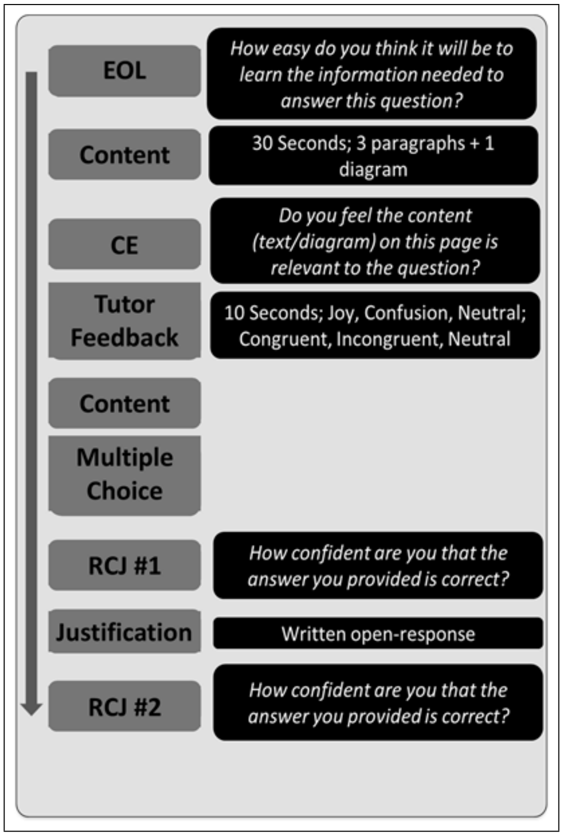

Figure 1: MetaTutor-IVH Experimental Paradigm

choice question about the content they just learned, they were asked to answer *"How confident are you that the answer you provided is correct?"*on a scale from 50 to 100 in which a score of 50 indicated the participant felt they had simply guessed. This acted as the immediate RCJ (RCJ1). They were then required to provide a justification for their answer in a text-based free response section. A second delayed RCJ (RCJ2) was asked once participants provided this reflection.

Experimental Design

The study used a 3x3x2 within-subjects design. First, trials had varying relevance levels: fully relevant (the text and diagram both contained pertinent information), text somewhat relevant (the diagram provides the pertinent information), diagram somewhat relevant (the text provides the pertinent information). Second, trials had three possible agent facial expressions presented: neutral facial expression (no change to the baseline agent expression), congruent facial expression (the agent expresses it is happy when the content is fully relevant and confused when the content is only somewhat relevant), and incongruent facial expression (the agent expresses it is confused when the content is fully relevant and happy when the content is only somewhat relevant). Finally, two types of questions were asked about the human body: a body function (e.g. *"Please explain how cortisol travels in the body."*) and a body malfunction (e.g., *"Please explain what would happen if the thyroid hormone were to diffuse freely from the thyroid all the time."*) . The instructional content always provided enough information to answer each question, however the source of the pertinent information differed.

Experimental Procedure

Participants were calibrated to a wireless Shimmer 3+ electrodermal activity bracelet, eye tracker, and affect recognition software before completing a demographic questionnaire, questionnaires gauging emotion and motivation followed by a biology content multiple-choice pretest. After the pretest, participants completed 18 trials in MetaTutorIVH. Following the trials, participants answered a series of additional motivation and emotion questionnaires, were debriefed, compensated, and thanked for their participation.

Apparatuses

During the 18 trials, data was collected from eye movements, emotion recognition software, EDA, and log files of the learners' interactions. Eye movements (not used in this study) were captured with an SMI RED 250 with a 60 Hz sampling rate (downshifted from 250 Hz to allow integration with iMotions Attention Tool). Affect (not used in this study) was captured using a web camera before being detected and automatically coded by iMotions FACET. EDA (not used in this study) was captured with a Shimmer 3+ wireless bracelet with a 128 Hz sampling rate. Finally, all log files and data streams were collected and aligned with iMotions Attention Tool 6.2 software (iMotions, 2016).

Coding and Scoring

Ease of Learning (EOL) Judgements To assess the precision of the judgment, we calculated the absolute accuracy index (AAI; (Schraw, 2009)). It is important to highlight that because this score is the discrepancy between judgment and performance, smaller scores correspond to higher accuracy.

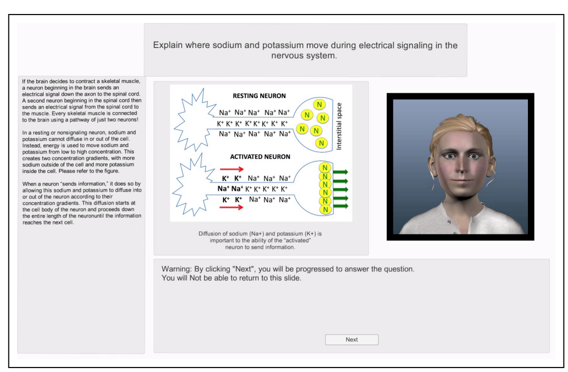

Figure 2: Sample MetaTutor-IVH Content Slide

Table 1: Correlations of all metacognitive absolute accuracy index

	M	SD	1	2	3
1. EOL	0.016	0.004			
2. CE	0.016	0.005	0.270*		
3. RCJ1	0.013	0.005	0.394*	0.449**	
4. RCJ2	0.014	0.004	0.354*	0.532**	0.803**

*Note: * indicates p <0.05; ** indicates p <0.005. EOl: Ease of learning; CE: Content Evaluations; RCJ1/2: Immediate/delayed retrospective confidence judgment*

Content Evaluations (CEs) A completely accurate CE was awarded .5 point (up to 1 point per trial), while a partially correct CE was awarded 0.25 points. An incorrect CE did not earn the participant any points. For example, in a trial in which the diagram was somewhat relevant and the text was fully relevant, if a participant reported both the text and diagram as fully relevant, they earned 0.75 points for that trial (0.5 for the diagram's completely correct judgment and 0.25 for the text partially correct judgment). Like EOLs, we calculated CEs AAI according to Schraw (2009).

Retrospective Confidence Judgements (RCJs) RCJ AAI was also calculated according to Schraw (2009). As with the other accuracies, the higher the absolute difference, the less accurate the participant's judgment was.

Equally-Weighted Composite Metacognitive Judgement Score We calculated an equally weighted composite score that consisted of all four metacognitive monitoring judgments accuracies.

$$Composite\ Score = \frac{EOL + CE + RCJ1 + RCJ2}{4} \quad (1)$$

Results

Preliminary

Participants (N = 55) on average answered 62% (SD = 0.14) of trials correctly (about 11/18 trials). All metacognitive judgments AAI were correlated with one another (see Table 1). Due to this multicollinarity, a multivariable statistical model for our second and third research questions was deemed inappropriate.

Which metacognitive judgment best captures performance variation during learning with MetaTutor-IVH?

Using each metacognitive judgment accuracy (i.e., EOL, CE, immediate RCJ, reflection RCJ) as a single explanatory variable, we ran four simple linear regression models to predict participant performance on the multiple-choice questions embedded in MetaTutor-IVH (see Table 2 for model's parameters, F, p, and R^2 statistics). EOL's AAI was a significant predictor of performance (F = 8.716 (1,54), p = 0.005), suggesting that for every 0.01 increase in the index value (i.e., the less accurate the participant was), performance decreased by .139 points (about 2.5 additional incorrect trials). CE's AAI was a significant predictor of performance (F = 45.18 (1,54), p <0.0005), suggesting that for every 0.01 increase in the index value (i.e., the less accurate the participant was), performance decreased by .205 points (about 3.7 additional incorrect trials). RCJ1's AAI (Immediate RCJ) was a significant predictor of performance (F = 8.2 (1,54), p = 0.006), suggesting that for every 0.01 increase in the index value (i.e., the less accurate the participant was), performance decreased by .137 points (about 2.5 additional incorrect trials). Finally, RCJ2's AAI (delayed) was also a significant predictor of performance (F = 24.1 (1,54), p <0.0005), suggesting that for every 0.01 increase in the index value (i.e., the less accurate the participant was), performance decreased by .216 points (about 3.9 additional incorrect trials). CE AAI was able to explain 47.47% of the variance in participant's performance, making it the strongest explanatory variable of all of the judgments. RCJ2 AAI's performed the next best, explaining 32.62% of performance variance. RCJ1 and EOL AAI both performed about the same, explaining 14.09% and 14.85%.

How does a model using a composite score compare to models only using individual components?

Using the uniformly weighted composite score (see Equation 1), we ran a simple linear regression to model participant performance. This score was a significant predictor (F = 39.9 (1,54), p <0.0005), and suggested that for every 0.01 increase in the score, performance decreased by 0.300 points (about 5.4 additional incorrect trials). This model was able to explain 44.39% of the variation in performance. Compared to the previous models, it outperformed all except for the CE AAI.

How does an equally weighted composite score model compare to a model weighted by our findings in (RQ2)?

We modified our original composite score to reflect how well each judgment was able to explain performance variation. Because there is currently no empirical or theoretical evidence for how all judgments should be weighted, we naively ranked with a 10% increase in weight per variable ranked by their R^2s (see Equation 2).

$$\begin{aligned} Composite\ Score = (0.4 * EOL) + (0.3 * CE) + \\ (0.2 * RCJ1) + (0.1 * RCJ2) \end{aligned} \quad (2)$$

Table 2: Model parameters, F, p, and R^2

	Intercept	B	R^2	F (1,54)	p
EOL	0.839	-13.917	0.149	8.716	<0.005
CE	0.955	-20.531	0.475	45.181	<0.005
RCJ1	0.802	-13.742	0.141	8.200	0.006
RCJ2	0.915	-21.647	0.326	24.100	<0.005

This new score was a significant predictor of performance (F = 55.48 (1,54), $p < 0.0005$), suggesting that for every 0.01 increase in the score (i.e., as accuracy decreased) we would expect a decrease in 0.311 points (about 5.6 additional incorrect trials). This new model was able to explain 52.60% of the variation in performance which outperformed all previous models.

Discussion

To explore how different metacognitive monitoring judgments taken throughout the entire learning session could be used in conjunction to better capture performance variation, we compared six models of participants using a multimedia learning environment while they learned about complex human biology systems. We showed that even with a naively weighted composite score accounting for temporal fluctuations of metacognitive monitoring accuracy, we were able to better capture performance variation than any singular judgment could capture alone. The first four models we used were all able to significantly describe the variation of performance to varying degrees of success, which is consistent with previous research (Dever et al., 2019; Dougherty et al., 2005; Hines et al., 2009; Jemstedt et al., 2017; Pilegard & Mayer, 2015; Ryals et al., 2016). Similarly, we found that CEs were the best predictor of performance while EOLs performed the worst, supporting our hypothesis. Furthermore, delayed RCJs outperformed immediate RCJs which suggest that reflection can increase the accuracy of metacognitive monitoring processes (Efklides, Schwartz, & Brown, 2018). This also provides support to the idea that judgments occurring later in the learning session are influenced by previous judgments. CEs proved to be the strongest predictor, which we believe is due to the direct allocation of effort and attention. While outside the scope of this analysis, it is also possible that the agent's given CE could have also effected the accuracy of both RCJs, and potentially in different ways. Future analysis will explore any differences in both the accuracy of these judgements, and behaviors of the participants once a judgement was provided by the agent. When accounting for these differences in our weighted composite score (RQ3), we were able to create a statistical model that captured 52.60% of performance variation. More statistically sophisticated weighting (i.e., PCA or factor analysis) might provide stronger weight values that could increase model performance even more. However, first more research should examine the relationship of these metacognitive monitoring judgments across multiple tasks across various domains to see if they remain consistent. For example, do CEs always provide the most information, or only for text-heavy environments? If similar patterns emerge, we might conclude that judgments occurring later in the learning session are affected by previous judgments and decisions based on those judgments. Future research should also consider capturing metacognitive monitoring through other measures outside of self-reports such as eye-tracking or think-aloud protocols. These measures would allow for even more granular analysis of the temporal fluctuations of metacognitive monitoring processes.

Metacognitive monitoring is a dynamic process, and empirical models aimed at capturing those processes should reflect this. Judgments that are made later in a learning session appear to be influenced by prior judgments and the experience gained from decisions based on monitoring. This study has begun to explore ways that we can naively capture the variation of performance by considering multiple judgments and considering their relationship to one another. Content evaluations, which occur during learning, are the strongest predictors of performance while ease of learning judgments are poor predictors. Retrospective confidence judgments were also strong predictors, especially after prompted reflection of the learned material. A model that accounts for these differences was able to outperform models that looked at the processes independently of one another. Future research on metacognitive monitoring should begin by validating and replicating similar findings within new environments in order to explore how how different judgments effect one another and what processes are shared across the learning session and during different tasks, domains, and learning techniques.

References

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self-and socially-regulated learning. *Metacognition and Learning*, *9*(2), 217–228.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of educational psychology*, *96*(3), 523.

Azevedo, R., Mudrick, N. V., Taub, M., & Bradbury, A. (2019). Self-regulation in computer-assisted learning systems. In J. Dunlosky & K. A. Rawson (Eds.), *The cambridge handbook of cognition and education* (chap. 23). Cambridge University Press.

Azevedo, R., Taub, M., & Mudrick, N. V. (2018). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In J. A. Greene & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 254–270). NY, New York: Routledge.

Bannert, M., Hildebrand, M., & Mengelkamp, C. (2009). Effects of a metacognitive support device in learning environments. *Computers in human behavior*, *25*(4), 829–835.

Brewer, W. F., & Sampaio, C. (2012). The metamemory approach to confidence: A test using semantic memory. *Journal of Memory and Language*, *67*(1), 59–77.

Britton, B. K., Van Dusen, L., Gülgöz, S., Glynn, S. M., & Sharp, L. (1991). Accuracy of learnability judgments for instructional texts. *Journal of Educational Psychology*, *83*(1), 43.

Chow, P. I., & Berenbaum, H. (2012). Perceived utility of emotion: The structure and construct validity of the perceived affect utility scale in a cross-ethnic sample. *Cultural Diversity and Ethnic Minority Psychology*, *18*(1), 55.

Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: neural correlates of the cognitive process and subjective level of confidence in recognition memory. *Neuroimage*, *29*(4), 1150–1160.

Chua, E. F., & Solinger, L. A. (2015). Building metamemorial knowledge over time: insights from eye tracking about the bases of feeling-of-knowing and confidence judgments. *Frontiers in psychology*, *6*, 1206.

de Boer, H., Kostons, A. S., & M., D.-B. D. D. N. (2012). *Effective Strategies for Self-regulated Learning : A Meta-Analysis* (Unpublished doctoral dissertation). Gronings Instituut voor Onderzoek van Onderwijs,.

Dever, D. A., Wiedbusch, M., & Azevedo, R. (2019). Learners' gaze behaviors and metacognitive judgments with an agent-based multimedia environment. In *International conference on artificial intelligence in education* (pp. 58–61).

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students . A meta-analysis on intervention studies at primary and secondary school level. *Metacognition Learning*(3), 231–264.

Dougherty, M. R., Robey, A. M., & Buttaccio, D. (2018). Do metacognitive judgments alter memory performance beyond the benefits of retrieval practice? a comment on and replication attempt of dougherty, scheck, nelson, and narens (2005). *Memory & cognition*, *46*(4), 558–565.

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & cognition*, *33*(6), 1096–1115.

Dunlosky, J., & Rawson, K. A. (2019). *The cambridge handbook of cognition and education*. Cambridge University Press.

Efklides, A., Schwartz, B. L., & Brown, V. (2018). Motivation and affect in self-regulated learning: Does metacognition play a role?

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin ' s model of self-regulated learning : New perspectives and directions. *Review of Educational Research*, *77*(3), 334–372.

Gross, J. J., & John, O. P. (2003). Individual Differences in two emotion regulation processes : Implications for affect , relationships , and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348–362.

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, *24*(2), 462.

iMotions. (2016). *Attention tool (version 6.2).* apparatus and software. Cambridge,MA.

Jansen, R. S., Leeuwen, A. V., Janssen, J., & Jak, S. (2019). Self-regulated learning partially mediates the effect of self- regulated learning interventions on achievement in higher education : A meta-analysis. *Educational Research Review*, *28*(September), 1–20.

Jemstedt, A., Kubik, V., & Jönsson, F. U. (2017). What moderates the accuracy of ease of learning judgments? *Metacognition and Learning*, *12*(3), 337–355.

Jönsson, F. U., & Kerimi, N. (2011). An investigation of students' knowledge of the delayed judgements of learning effect. *Journal of Cognitive Psychology*, *23*(3), 358–373.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of experimental psychology: Learning, Memory, and Cognition*, *16*(3), 464.

Mayer, R. E. (2014). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. *The Cambridge handbook of multimedia learning*, *16*, 345–370.

Mazzoni, G., Cornoldi, C., Tomat, L., & Vecchi, T. (1997). Remembering the grocery shopping list: A study on metacognitive biases. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *11*(3), 253–267.

McCarley, J. S., & Gosney, J. (2005). Metacognitive judgments in a simulated luggage screening task. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 49, pp. 1620–1624).

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The psychology of learning and motivation*, *26*, 125–141.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (aeq). *Contemporary Educational Psychology*, *36*, 36–48.

Pilegard, C., & Mayer, R. E. (2015). Adding judgments of understanding to the metacognitive toolbox . *Learning and Individual Differences*, *41*, 62–72. doi: 10.1016/j.lindif.2015.07.002

Pintrich, P. R., Elder, A. D., Garcia, T., Anderman, L. H., Hofer, B., Patrick, H., ... Urdan, T. (2000). Multiple goals , multiple pathways : The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, *92*(3), 544–555.

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*(1), 2–13.

Ryals, A. J., Rogers, L. M., Gross, E. Z., Polnaszek, K. L., & Voss, J. L. (2016). Associative recognition memory awareness improved by theta-burst stimulation of frontopolar cortex. *Cerebral Cortex*, *26*(3), 1200–1210.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and learning*, *4*(1), 33–45.

Schunk, D. H., & Greene, J. A. (2017). *Handbook of self-regulation of learning and performance*. Routledge.

Serra, M. J., & Ariel, R. (2014). People use the memory for past-test heuristic as an explicit cue for judgments of learning. *Memory & cognition*, *42*(8), 1260–1272.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1379.

Taub, M., & Azevedo, R. (2018). Using sequence mining to analyze metacognitive monitoring and scientific inquiry based on levels of efficiency and emotions during game-based learning. *JEDM— Journal of Educational Data Mining*, *10*(3), 1–26.

Tauber, S. K., & Rhodes, M. G. (2012). Multiple bases for young and older adults' judgments of learning in multitrial learning. *Psychology and aging*, *27*(2), 474.

Usher, E. L., & Schunk, D. H. (2018). Social cognitive theoretical perspective of self-regulation.

Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. In J. A. Greene & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 52–64). NY, New York: Routledge.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. *Metacognition in educational theory and practice*, 277–304.

Winne, P. H., & Hadwin, A. F. (2012). The weave of motivation and self-regulated learning. In *Motivation and self-regulated learning* (pp. 309–326). Routledge.