

Assessing Metacognitive Components in Self-Regulated Reading of Science Texts in E-Based Environments

Miao-Hsuan Yen¹ · Chia-Yu Wang²  ·
Wen-Hua Chang¹ · Sufen Chen³ · Ying-Shao Hsu¹ ·
Tzu-Chien Liu⁴

Received: 14 August 2016 / Accepted: 27 March 2017 / Published online: 11 April 2017
© Ministry of Science and Technology, Taiwan 2017

Abstract The boom in Internet and communication technology has accelerated change in information delivery channels from paper to networked, digital displaying devices. Reading electronically has altered the cognitive and metacognitive processes readers use to approach science texts. Assessing reading in a digitized environment also opens

Electronic supplementary material The online version of this article (doi:10.1007/s10763-017-9818-2) contains supplementary material, which is available to authorized users.

✉ Chia-Yu Wang
cywang@g2.nctu.edu.tw

Miao-Hsuan Yen
myen@ntnu.edu.tw

Wen-Hua Chang
sujudy@ntnu.edu.tw

Sufen Chen
sufchen@mail.ntust.edu.tw

Ying-Shao Hsu
yshsu@ntnu.edu.tw

Tzu-Chien Liu
tzuchien@ntnu.edu.tw

- ¹ Graduate Institute of Science Education, National Taiwan Normal University, 88, Sec. 4, Ting-Jou Rd, Taipei City 116, Taiwan
- ² Institute of Education, National Chiao Tung University, 1001, Ta-Hsueh Rd, Hsinchu City 300, Taiwan
- ³ Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, 43, Sec. 4, Keelung Rd, Taipei City 106, Taiwan
- ⁴ Department of Educational Psychology and Counseling, National Taiwan Normal University, 162, Sec. 1, Heping E. Rd, Taipei City 106, Taiwan

up opportunities to examine the recursive process of self-regulated learning. However, a systematic analysis in this area across paper- and e-based environments is lacking. This study explores methods of assessing metacognition in reading informational science texts to address this gap. A total of 55 studies in 47 articles published from 1990 to 2016 investigating effects of metacognition on science reading are reviewed. The analysis reveals frequently applied methods, including self-report questionnaires for measuring metacognitive knowledge, event-based assessment for metacognitive skills, and questionnaires or interviews for metacognitive experiences. Community acceptable instruments are presented and illustrated. E-based environment has potential to support comprehensive and complicated measurement of metacognition. Future research may further the investigation of interactions among metacognitive components by designing sequential tasks and combining multiple assessment methods. Finally, triangulation of findings from various methods is essential in assessing metacognition involved in self-regulated reading.

Keywords Metacognitive assessments · Science reading · Self-regulated learning

Introduction

Reading science texts is a major practice of science since many ideas and results of inquiry are communicated and disseminated in writing. Science texts, whether in textbooks, media, or webpages, consist of science facts, principles, theories, or procedures related to science research activities. Literate citizens must read reports about science in the press or on the Internet and understand the salient concepts. They also need to discuss the validity and reliability of the data and conclusions to make proper daily decisions and participate in social scientific issues. The boom in Internet use and the widespread use of learning technology have opened up new channels and opportunities for learning. For instance, open online repositories of educational materials, online courses, and hypermedia environments have made abundant resources available for free access. Knowing how to seek and use such massive informational resources demands more learner control of self-directed learning. Students need sustained practice and support to develop the ability to extract and construct meaning from science texts in books, media reports, and other forms of scientific communication. Fostering scientific literate, self-regulated learners is more urgent in this information explosion era than ever before.

The change in the information delivery channel from paper to networked, digital display devices (i.e., e-based environment or e-based reading) has expanded the definition of reading to encompass both static and dynamic texts, and has altered the mental processes that readers use to approach texts (Michalsky, 2013; Organization for Economic Co-operation and Development [OECD], 2013). Reading and understanding of science texts in an electronic environment have been creating new demands of reading proficiency, whereby learners have to construct meaning through supplementing tools, cues/guidance, or multimedia. Effective science reading involves not only searching, analyzing, and integrating text and multimedia (including diagrams, animations, videos, and mathematical representations) information retrieved from a large amount of material but also metacognitively setting learning goals, monitoring strategy use and state of

understanding, and critically judging information credibility (Azevedo & Cromley, 2004; Keck, Kammerer & Starauschek, 2015; Mason, Junyent & Tornatora, 2014). Reading in such an environment with a high degree of freedom can be challenging if students do not possess adequate metacognitive skills (Yee, Hsieh-Yee, Pierce, Grome & Schantz, 2004). Science teachers need to understand what facets of metacognition are associated with better science reading performance to provide appropriate metacognitive scaffolds. Instructors and curriculum developers also need knowledge of metacognitive assessment to assess and interpret what students know and are able to do, in order to design proper instructions or supporting tools to enhance science reading competency, regardless of whether it takes place in a paper- or an e-based reading situation. Through conducting a systematic discussion on methods and tools for measuring metacognition, we hope to generate useful insights and suggestions for designing and implementing appropriate and adequate assessments. This is important because appropriate assessment would yield high-quality evidence for teachers and researchers to explore critical components of learning and to evaluate the effectiveness of instruction.

A few reviews have been carried out to analyze and discuss the overall trends of research on metacognition in learning (Azevedo, 2009; Donker, de Boer, Kostons, Dignath van Ewijk & van der Werf, 2014), in science education (Zohar & Barzilai, 2013), in e-based learning environments (Winters, Greene & Costich, 2008), and in science reading (Hsu, Yen, Chang, Wang & Chen, 2016). These reviews have in part examined issues in measuring and instructing metacognition in general (Akturk & Sahin, 2011), in learning strategy instruction (Donker et al., 2014), during hypermedia learning (Azevedo, Moos, Johnson & Chauncey, 2010; Devolder, van Braak & Tondeur, 2010), and reading in general (e.g., Jacobs & Paris, 1987). Yet, a systematic analysis that explores issues and challenges of measuring metacognition in science reading literature has not been documented in previous reviews. To address the above-mentioned gaps, the purpose of this study is to review and synthesize methods for assessing metacognition in science reading. Recent methodological issues and the advantages/constraints of various methods for assessing metacognitive components are then discussed with consideration of the characteristics of reading environments. By doing so, we hope to provide some insights and suggestions for education professionals, including science teachers and researchers, to design quality metacognitive assessments, and to use the results of these measurement approaches.

Theoretical Framework

Imagine that a teacher gives a homework assignment asking students to write a report on “how my actions contribute to global climate change” after they have learned about the natural greenhouse effect. Each group explores and studies some preselected websites about fuel, electricity, and carbon dioxide to develop their arguments on whether they or their parents should choose an electric car, an alternative-fuel vehicle, or a gasoline car. While navigating through the websites, effective learners analyze the learning situation and make decisions on what to read, how to read it, and how much to read to support their standpoint and to refute counter stances. They must set meaningful learning goals and plan which strategies to use. When processing texts, it is critical that

they are able to make accurate judgments on which parts of informational text have been understood well and which have not. This requires the students to constantly monitor their ongoing learning progress. They also need to reflect on their learning outcomes to adjust their goals, plans, strategies, or efforts accordingly. Collectively, the reading activities involve metacognitive monitoring and control, and are sometimes also called self-regulated learning (SRL).

Reading scientific texts along with hypermedia and/or supporting tools of learning technologies can be very taxing. Failure in self-regulation during science reading may be attributed to some students' lack of appropriate metacognitive knowledge (MK). It is also possible that the students possess the requisite MK but do not know when and how to deploy it in a specific context (in other words, they do not have adequate metacognitive skills (MSs) [e.g., Bannert & Mengelkamp, 2013]). Even when students hold certain MK and are able to demonstrate MS appropriately, they may misjudge their level of understanding due to an inadequate metacognitive experience (ME) and, in turn, terminate their efforts prematurely (e.g., Lauterman & Ackerman, 2014).

Several frameworks or models have been used to account for students' metacognitive and self-regulated processes in science reading, for example, Pintrich's (2000) framework of SRL as well as Zimmerman's (2000) and Winne and Hadwin's (1998) models. According to these models, self-regulated learning unfolds in three or four possible, sequenced phases, including task definition, goal setting and planning, enacting study strategies and monitoring, and evaluation. We selected Winne and Hadwin's (1998) SRL model to conceptualize how enactment and measurement of MK, MS, and ME influence the learning process and outcomes of science reading because it describes self-regulation in great detail. However, a sophisticated model of learning needs to be supported with solid evidence yielded from high-quality assessment information. Measurement methods and tools are needed for eliciting and interpreting appropriate and relevant data from students. Thus, we bring in the definitions of three facets, MK, MS, and ME, from metacognitive theories to explore methods and tools for assessing metacognition in science reading literature. These definitions are further elaborated in the following sections.

Among metacognitive theories (e.g., Brown, 1987; Schraw & Dennison, 1994) and studies that explore influence of metacognition in learning, a distinction is often made between knowledge and regulation of cognition. Knowledge of cognition, so often called as MK, refers to declarative knowledge about the interplay between person, task, and strategy characteristics (Veenman, 2013). An individual realizes a task and makes proper control decisions within this framework of metacognitive knowledge (Efklides, 2006). MS entails that a learner consciously and purposively applies strategies to regulate his or her thinking or learning process to ensure the desired outcomes. In addition to measuring the amount and quality of MK and MS that learners possess or deploy, feeling of knowledge and judgment of learning have been investigated (Veenman, van Hout-Wolters & Afflerbach, 2006) as an additional, critical facet of metacognition. These are metacognitive experiences (ME) which involve online monitoring of cognition as people come across a task- and process-related information (Efklides, 2006). Accurate monitoring of learning experience is crucial for effective SRL. The influence of ME on learning has received increasing attention (e.g., a special issue in *Learning and Instruction*, vol. 24, 2013) because it may serve as feedback for regulating MS and, therefore, influences learning outcomes.

In Winne and Hadwin's (1998) model, the amount of MK can be captured and assessed as a factor of learners' cognitive conditions, and quality of MS can be measured as frequencies and types of metacognitive strategies activated during the ongoing reading task. In addition, their model portrays metacognitive monitoring as comparing product of studying with a set of internal criteria and generates an evaluation that serves as feedback at several learning phases. This provides a ground for portraying status, quality, and accuracy of ME, as well as their role on influencing SRL. The present study specifically addresses the following research questions:

1. What are the shared methods for assessing metacognitive process in science reading including MK, MS, and ME?
2. How do the measurements of MK, MS, and ME differ in paper- and e-based reading environment?

Method

A content analysis approach was adopted to summarize and synthesize assessment tools for metacognitive components used in the literature. The research procedure involved an initial search and screening for related empirical articles and cycles of coding and discussion.

Analysis Procedure

In conducting this review, we searched the Web of Science database for journal articles included in the Social Science Citation Index. The searching criteria were as used in our previous study (Hsu et al., 2016). In the first selection phase, three sets of keywords were combined to search for empirical studies in the area of educational research from 1990 up to 2016. These three sets of keywords addressed research areas regarding self-regulated learning ("metacogniti*" or "self-regulat*" or "metacompreh*"), reading ("text" or "hypertext"), and domains of science (scien* or physic* or chem* or bio* or engine* or earth* or geo* or medic*). We excluded articles addressing listening, disability, or involving patients (listen* or disab* or patient) since the approach and the participants of the studies were outside the focus of the present study. The search resulted in 162 hits.

In our second selection phase, we went through the title, abstracts, introductory and methodological sections of the found studies and selected those that met the following criteria: (1) The articles were empirical studies, (2) the variable "metacognition" or "self-regulation" was described and measured, and (3) the context involved reading or text processing in a science domain, excluding mathematics and engineering. Studies which implemented metacognitive training without measuring metacognitive components before, during, and after instruction were excluded from the analysis. In addition, studies which measured epistemic metacognition and general self-efficacy were also excluded. In the end, 47 articles (55 studies) entered the final review.

Coding Scheme

The coding scheme is displayed in Table 1. Regarding the characteristics of the reading tasks, two codes for *mode* (delivery channel) and *text display space* were adopted from the draft 2015 PISA reading framework to depict the characteristics of the task design. Regarding mode, text can be delivered on paper (paper-based) or on a screen, including computers or electronic devices such as a smart phone or a tablet (e-based). Texts that were embedded in test items for assessing metacognitive knowledge of science text and reading (Yore, Craig & Maguire, 1998) or for rating expected performance (Linderholm, Zhao, Therriault & Cordell-McNulty, 2008) were categorized as “others.” Text display space consists of three subcategories: static, dynamic text, and others. Examples of “static text” include paragraphs extracted from science textbooks (e.g., Michalsky, Mevarech & Haibi, 2009), articles in science magazines (e.g., Chularut & DeBacker, 2004), and journal articles (e.g., Dhieb-Henia, 2003). Tasks designed for assessing SRL with dynamic texts included searching for information related to a specific topic using a search engine (e.g., Coiro & Dobler, 2007; Tu, Shih & Tsai, 2008) and reading hypertext or hypermedia material for comprehension (e.g., Greene, Moos, Azevedo & Winters, 2008) or for essay revision (Butcher & Sumner, 2011). Conducting online inquiry (Raes, Schellens, De Weyer & Vanderhoven, 2012) involves navigating information across multiple webpages and was categorized into dynamic text. Others include writing learning journals to document strategies used when studying the textbook as homework (e.g., Glogger, Schwonke, Holzäpfel, Nückles & Renkl, 2012) or using a computer environment for training technological pedagogical content knowledge of preservice teachers (Kramarski & Michalsky, 2010). As shown in Table 2, the combination of *mode* and *text display space* results in four categories including paper-static, e-static, e-dynamic, and others. No study investigated reading dynamic text on paper.

The *components of metacognition*, as addressed in the theoretical framework, include MK, MS, and ME. MK is declarative knowledge of cognition which comprises knowledge of oneself as a learner, different types of tasks, and their processes. MK also involves a repertoire of strategic knowledge, including planning, monitoring, and evaluation, for achieving a particular task (Schraw, 1998). MSs depict skills that one purposefully utilizes to control cognition. Similar to MK, MS generally comprises three

Table 1 Coding scheme

Category	Associated codes
Characteristics of tasks	
Mode	Paper-based, e-based, others
Text display space	Static text, dynamic text, others
Components of metacognition	Metacognitive knowledge Knowledge of cognition (declarative, procedural, conditional) Regulation of cognition (planning, monitoring, evaluation, etc.) Metacognitive skills (planning, monitoring, evaluation, etc.) Metacognitive experience (judgment of learning, feeling of confidence, effort needed, and calibration)
Types of assessment	Questionnaire, interview, think-aloud, observation, computer logs, others

Table 2 Summaries of metacognitive components, types of assessment, and media

Metacognitive components ^a	Types of assessment ^a	Paper-static (<i>N</i> = 21)	E-Based		Others (<i>N</i> = 4)	Subtotal
			Static (<i>N</i> = 15)	Dynamic (<i>N</i> = 15)		
MK (<i>n</i> = 10)	Questionnaire	7	0	1	2	10
MS (<i>n</i> = 30)	Questionnaire	1	0	1	0	2
	Interview	4	1	1	0	6
	Think aloud	5	1	7	0	13
	Observation	0	0	1	0	1
	Computer logs	0	0	4	0	4
	Others	0	0	2	2	4
ME (<i>n</i> = 23)	Questionnaire	5	14	1	1	21
	Interview	2	0	0	0	2

^a Some studies may include more than one type of assessment

MK metacognitive knowledge, *MS* metacognitive skills, *ME* metacognitive experience, *N* = numbers of studies

subcategories: planning, monitoring, and evaluating. Planning involves comprehending the task requirements, and selecting appropriate strategies as well as allocating resources accordingly for task processing (Schraw, 1998; Veenman & Elshout, 1999). Monitoring refers to continuously overseeing cognitive processes and regulating them when they fail, whereas evaluating indicates checking if the outcome is in line with the goal (Veenman & Elshout, 1999). Although MK and MS share some subcategories and definitions, MK emphasizes knowledge and experience of using these strategies that is stored in long-term memory. MS, on the other hand, reflects real-time actions of these strategies executed to regulate cognition on the fly. ME includes metacognitive feelings (Koriat & Levy-Sadot, 2000) and judgments regarding learning process and performance. They are typically measured by judgments/estimates of learning (Dinsmore & Parkinson, 2013) and feelings of confidence, or how much effort is needed, etc.

Finally, *assessment types* used for detecting metacognition were categorized into questionnaires, interviews, think-aloud, observations, computer logs, and others. The coding work was conducted by the authors, who are science educators with research experience in metacognition and have conducted a review of science reading (Hsu et al., 2016). We randomly selected 26 articles for the first run of analysis. Each article was analyzed by two coders, and the interrater agreement was 0.81. Discrepancy was resolved through discussion (with a third person when necessary). The remaining articles were independently coded by one of the authors.

Results

The assessment types used to measure each metacognitive component in the two delivery media are summarized in Table 2. Details of the characteristics of each study are provided in the Appendix. Among the 55 reviewed studies, 21 were implemented with static text in a paper-based environment, 15 with fixed and static reading material

(e.g., on PowerPoint slides or Word) using electronic devices, and 15 with dynamic text. Overall, the majority of them ($n = 30$) assessed learners' metacognitive skills. Ten and 23 studies measured metacognitive knowledge and experience, respectively. In the following sections, for Research Question 1, we analyze features of the assessments in MK, MS, and ME, illustrated with technique and examples for task designs. Subcodings under MK, MS, and ME were extracted from the reviewed literature. Concerning Research Question 2, methodological issues for assessing particular metacognitive components were then discussed with consideration of features of e-based reading environments.

Measuring Metacognitive Knowledge

MK portrays the extent to which an individual possesses knowledge of cognition concerning self as a learner, tasks, and strategies. Accumulated from multiple learning episodes in the past, MK informs how well the learner performed on a particular task with specific strategies. Such information could lead the learner to adopt or improve a strategy in a similar episode. Thus, MK might be the repertoire of MS.

In the reviewed articles, MK was always assessed by self-report questionnaires, mostly (7 out of 10 articles) with 3–5-point Likert scales or a continuous 100-mm bipolar scale. The text anchors at the two extremes were never/always or agree/not agree. These questionnaires were often used to examine effects of an instruction in a pre- and posttest setting, and the pretest MK could be treated as learners' characteristics to explore its influence on training effects. Questionnaires assessing MK generally include one of the two subscales, namely "knowledge of cognition" and "regulation of cognition" (e.g., Metacognitive Awareness Inventory; Schraw & Dennison, 1994). Items in the "knowledge of cognition" subscale probe for knowledge about person, tasks, and strategies in categories such as declarative, procedural, and conditional knowledge. Items of "regulation of cognition", on the other hand, assess participants' perceived usage of planning, monitoring, evaluation, and implementation of strategies. Examples of survey items are shown in Table 3.

Item descriptions could be altered for assessing different MK components in general or task-specific conditions. For example, an original item in Metacognitive Awareness Inventory that assesses general declarative knowledge such as "I am good at organizing information" could be modified as "I was good at presenting the information I had found on the Internet" to grasp task features of web-based inquiry. Based on our reviewed studies, revising a general MK questionnaire such as Metacognitive Awareness Inventory or the Motivated Strategies for Learning Questionnaire (Pintrich, Smith, Garcia & McKeachie, 1991) into a task-specific instrument (e.g., inquiry, Michalsky et al., 2009; problem-solving, Raes et al., 2012) is a common approach for conducting MK assessment. Since different tasks or reading environments demand different metacognitive behaviors (e.g., using keywords to search for information on the Internet and reading hypertext are different from reading static text), researchers interested in probing task-specific metacognitive knowledge and behavior should ensure that the items capture the specific behavior. However, it could be a challenge to foresee how participants might apply their knowledge to accomplish a "novel" task in a technology-rich environment, and design questionnaire items accordingly in advance.

Table 3 Examples of questionnaire items assessing metacognitive knowledge

Measurement	Sample questions
Knowledge of cognition	
Declarative knowledge	
Task	I know what kind of information is most important to learn. ^a (Schraw & Dennison, 1994)
Person	I was good at presenting the information I had found on the Internet. ^b (Raes et al., 2012)
Procedural knowledge	
Strategies	I have a specific purpose for each strategy I use. ^a (Schraw & Dennison, 1994)
Task	I plan the stages of performing the inquiry task. ^b (Michalsky et al., 2009)
Conditional knowledge	
Strategies	I use different learning strategies depending on the situation. ^a (Schraw & Dennison, 1994) During planning, I paused to consider if I know what to do. ^a (Michalsky et al., 2009)
Regulation of cognition	
Planning	In this class, I set goals for myself which I try to accomplish. ^a (García, García, Berbén, Pichardo & Justicia, 2014) When I begin to work on the task for the course, I think what would be a good way to do it. ^b (Kramarski & Michalsky, 2010) I plan the order of activities for solving the problem. ^{a or b} (Michalsky et al., 2009)
Monitoring	In this class, I try to monitor my progress when I study. ^a (Shell, Husman, Turner, Cliffel, Nath & Sweany, 2005) During the task process I often ask myself if I am going in the right direction. ^b (Kramarski & Michalsky, 2010) While searching the Internet for information, I often asked myself if my strategy would result in a good answer for the information problem. ^{a or b} (Raes et al., 2012) When I read a scientific text, I force myself to check to see if I remember what I have learned. ^a (Michalsky, 2013)
Evaluation	In this class, I try to determine the best approach for studying each assignment. ^a (Shell et al., 2005) At the end of the task, I ask questions to make sure I know the material I have been studying. ^b (Kramarski & Michalsky, 2010)

^a General^b Task-specific

Levels of MK can also be assessed in a multiple-choice test format. In Index of Science Reading Awareness, developed by Yore et al. (1998) and adapted in Wang and Chen (2014) and Wang, Chen, Fang, and Chou (2014), for instance, the multiple-choice items were written to target specific knowledge of science reading, science text, and science reading strategies. The choices were coded as comprehensive/an interactive-constructive interpretation of reading (2 points), surface/a top-down or bottom-up interpretation of reading (1 point), unrelated or trivial (0 points) (Wang et al., 2014). For a situation in which the test takers come to a word that they do not

know, the response “I skip over and read on” was coded as trivial and received 0 points, whereas “I use the words around it to think about its meaning” was considered as comprehensive and received the highest score.

The questions and timing for measuring MK are similar in paper- and e-based reading. They differ only in task-specific measures. One of the challenges is that participants may respond to the survey items by referring to their perceptions of how they performed on the task just executed or with an overall impression of what they have experienced in the whole training course. Furthermore, the timing of questionnaire implementation may also affect learners’ interpretation and responses of the items. For instance, if the questionnaire was distributed *immediately* after a task, it was likely that participants would respond to the items based on what they just experienced. If it was given at the end of a series of instructional activities, the basis might become the overall impression of the past events or the learning episode that was most available in memory. To increase data validity for assessing task-specific MK, researchers should word the task scenario carefully by avoiding ambiguous descriptions and make sure that participants understand which scenario is being examined.

In short, for measuring MK, paper- and e-based research uses similar questionnaires. Table 3 outlines the subconcepts of MK along with examples and references. For most of the subconcepts, both general and task-specific samples were extracted from the literature. Task-specific measures are recommended to couple with careful consideration of timing of administration.

Measuring Metacognitive Skills

In comparison to self-report MK measures, MS revealed by protocol analysis and tracing techniques reflects a dynamic, event-based view of metacognition. Methods like thinking-aloud, interviews, and recording log data are more capable of capturing not only what MS students possess but also when, how often, and in what process order these skills are deployed over a particular learning task (Greene, Robertson & Costa, 2011).

Among the reviewed studies (see Table 2), MS were assessed mostly using protocol analysis techniques. Thinking-aloud ($n = 13$) captures the utilization of MS by prompting participants to verbalize their thinking or decision-making processes while performing a task. Interviews, another common protocol analysis technique, often utilize a set of open-ended questions to guide participants to describe and give examples of their use of MS ($n = 6$). When implementing interviews, particular attention has been given to the likelihood of using particular strategies, and the amount and quality of MS they are aware of during the task or retrospectively after the task. Similar open-ended questions may be implemented in a written format to guide learners to articulate their use of MS in a particular task condition, such as completing reflection questions on a worksheet after reading and critiquing journal articles ($n = 1$, Dhieb-Henia, 2003), keeping a learning journal when doing homework ($n = 1$, e.g., Glogger et al., 2012), or sharing reflections in an online forum after attending a professional development program ($n = 1$, Kramarski & Michalsky, 2010). Examples of interview or written questions used to elicit different MS components are illustrated in Table 4.

The questions, coding scheme, and timing for measuring metacognitive skills are similar in paper- and e-based studying environments. To grasp event-based MS of a specific task aspect, prompting questions can be tailored to direct participants’

Table 4 Examples of interview or written questions assessing metacognitive skills

Measurement	Sample questions
Planning	<p>Do you have a goal (or goals) in mind as you prepare to do this activity? (Artino, Cleary, Dong, Hemmer & Durning, 2014)</p> <p>Do you have any particular plans for how to read this passage and take the test? (DiBenedetto & Zimmerman, 2013)</p> <p>Describe the goals of the task and explain how and why you set them prior to learning, referring to TPCK. (Kramarski & Michalsky, 2010)</p>
Monitoring	<p>As you have been going through this process, what has been the primary thing you have been thinking about or focusing on? (Artino et al., 2014)</p> <p>Can you remember any cues that helped you to supply the answers to the tasks? (Dhieb-Henia, 2003)</p> <p>When and how did you assess the TPCK activity, during analysis of the task? Please provide some examples. (Kramarski & Michalsky, 2010)</p>
Evaluation	<p>Describe the benefit you obtained from your cooperative learning process and how it helped you in advancing your TPCK learning task. Please provide some examples. (Kramarski & Michalsky, 2010)</p>

attention. For example, in Kramarski and Michalsky's (2010) study, preservice teachers' use of monitoring skills during the analysis and planning phases was elicited for the comprehending and designing aspects of a Technological Pedagogical Content Knowledge (TPCK) activity using separate questions. Moreover, providing guidance such as asking participants "Try to remember as accurately as possible how you carried out each of the tasks" (Dhieb-Henia, 2003) may help to ensure that learners report what they actually do rather than what they know or want to do.

Coding schemes for protocols may involve cognitive, metacognitive, and behavioral processes according to the focus of individual study. The number of codes used varied drastically across studies. Some studies analyzed fine-grained reading skills and utilized a coding scheme consisting of more than 30 codes in five categories: planning, monitoring, evaluation, strategy use, and task difficulty (e.g., Greene, Costa, Robertson, Pan & Deekens, 2010; Greene et al., 2008). There are also researchers who took a macro-analysis lens to examine fewer main skills. For instance, Franco, Muis, Kendeou, Ranellucci, Sampasivam and Wang (2012) categorized learners' actions including self-questioning, monitoring understanding, and acknowledging cognitive conflict, in order to investigate the influence of monitoring on conceptual change through reading a text presented in a reputational argument format. The other researchers had a special interest in specific monitoring skills such as monitoring understanding, monitoring environment, monitoring progress toward goals, monitoring use of strategies, and time monitoring (e.g., Moos & Azevedo, 2009).

In addition to the above-mentioned protocol analyses, tracing techniques using videos or computer logs offer other observable indicators as students engage with a task. For example, in Zhang and Quintana's (2012) study, a monitoring event was identified when students used a planning space embedded in the website to examine their progress and decide what should be done next. When reading and learning took place in a paper-based environment, a monitoring event was counted when students talked about their progress referring to their memory or their paper notebook. An evaluation event could also be recognized when students judged the accuracy of what they had already answered (reconfirmation) or when they modified (modification) their inputted responses after

making evaluations comparing their inputted answers with related information found on the web (She, Cheng, Li, Wang, Chiu, Lee, Chou et al., 2012).

Methodologically, there are advantages of assessing MS in e-learning environments in comparison with in paper-based environments. Computerized environments can engage dynamic texts, audio, video, and animated materials as well as search engines. This opens up a wide range of topics studying SRL in interaction with the rich content, learning platforms, and supporting tools. Unfortunately, relatively few ($n = 4$) among the reviewed e-based studies utilized tracing data. Nevertheless, these studies revealed that innovative measures of MS have been generated. For example, Tu et al. (2008) considered the number of keywords the participants used in search engines as an indicator of the ability of MS. Using fewer keywords was presumed to be stronger MS because this action requires better abilities of filtering information. Likewise, the frequency of using or viewing an embedded function or SRL tool demonstrates MS. For example, viewing the progress bar is monitoring learning, and writing goals or arranging a schedule in the provided space is planning. The features integrated in the learning platform could scaffold learning while also measuring MS. Finally, the adjustment of reading time on certain passages is associated with MS.

In sum, this section addresses methods for measuring MS using protocols and tracing data. Think-aloud and interview/written protocols are still the dominant methods. However, as noted by Winne and Perry (2000), incorporating data from observations or using the tracing technique is particularly important for assessing the metacognitive skills of young children who have limited language for articulating their reading processes. The tracing data may also be used to confirm or disconfirm metacognitive behaviors deployed in verbal or written protocols (Greene et al., 2011). By doing so, researchers are able to distinguish MS that students actually performed from what they claim to be capable of.

Measuring Metacognitive Experiences

Self-prediction of performance or judgment of comprehensibility while or after reading a text is an important metacognitive experience for learners to monitor how much effort is put into their learning. Metacognitive judgments are frequently used as alternative indicators to explore the effects of metacognitive regulation on reading (e.g., Burkett & Azevedo, 2012; Miles & Stine-Morrow, 2004). When making metacognitive judgments, students presumably infer how well they will do, either before or after task completion, based on certain information (e.g., the easiness of the task, or content familiarity) that is available to them.

Twenty-three reviewed studies in science reading have applied questionnaires ($n = 21$) and interviews ($n = 2$) to probe participants' ME, including judgment of the comprehensibility of a sentence or text, prediction of comprehension of a text, judgment of learning, and confidence. The questions used in questionnaires and interviews are similar, except that one is written and the other is oral. The questions and context of the measurements are illustrated in Table 5. Most of them are presented with Likert-type scales. Dichotomous scales (e.g., confident or not confident) are seldom adopted in studies of science reading. Judgment of learning or confidence is considered as a continuum over a spectrum.

Table 5 Examples of assessing metacognitive experience

Measurement	Sample questions and context
Ease of learning judgment	Before reading, the participants are presented with an open-ended inference question related to the topic they are going to read, and asked to rate “How easy do you think it will be to learn the information needed to answer this question?” on a 6-point scale (0 20% 40% 60% 80% 100%) (Burkett & Azevedo, 2012).
Comprehensibility rating	Participants rate the comprehensibility of each sentence on a 5-point scale from Very Poor (1) to Very Good (5). The rating is calibrated by the comprehensibility rating of the control sentence (Otero, Caldeira & Gomes, 2004). Participants rate the comprehensibility of each node they viewed in a hypertext on a 7-point scale from Very Comprehensible (1) to Very Incomprehensible (7) (Pieschl, Stahl & Bromme, 2008).
Judgment of sentence difficulty	Participants estimate the difficulty of the sentence just read on a 4-point scale (1 = very easy, 4 = very difficult) (Ozuru, Kurby & McNamara, 2012).
Immediate judgment of learning	Immediately after reading, participants rate “How well did you understand the information about the subject presented in the text you just read?” on a 6-point scale (0 20% 40% 60% 80% 100%) (Burkett & Azevedo, 2012). After reading, participants rate “How well do you think you understood the text?” on a 7-point scale (1 = very poorly, 7 = very well) (Ikeda & Kitagami, 2012).
Delayed judgment of learning	30 s after reading, participants rate “How well did you understand the information about the subject presented in the text you just read?” on a 6-point scale (0 20% 40% 60% 80% 100%) (Burkett & Azevedo, 2012).
Prediction of performance	“Now that you have described how you went about judging your learning of the texts you read, please estimate for each of the texts what grade you believe you might receive on a test over each text (e.g., A, B, C, D, E). Indicate the estimated grade in the space below.” (Linderholm et al., 2008). Participants drag an arrow along a continuous 25%–100% scale to indicate how well they thought they would do on (1) memory test questions and (2) higher-order comprehension test questions (Lauterman & Ackerman, 2014). Participants predict how well they thought they would answer a question based on the sentence just read on a 4-point scale (1 = likely to be wrong, 4 = likely to be correct) (Ozuru et al., 2012).
Confidence judgment on comprehension test	After the comprehension test, participants rate “How confident are you that the answer you provided is correct?” on a 6-point scale (50% 60% 70% 80% 90% 100%), where 50% was interpreted as guessing and 100% as completely confident (Burkett & Azevedo, 2012). Immediately following the response to each comprehension question, participants give a confidence rating on a 4-point scale (1 = not at all confident, 4 = very confident) (Ozuru et al., 2012). For each question, participants are asked “How confident do you feel about your answer to the question ...?” They rate on a 10-point scale ranging from not sure at all (10) to very sure (100) (DiBenedetto & Zimmerman, 2013).
Judgment base	“Now that you have read two texts and judged your level of learning of the texts, please describe in the space below how you made your judgments of learning for each text. That is, how did you go about deciding how well you learned from the texts? Please be as specific as possible when describing how you went about making your judgments” (Linderholm et al., 2008).

In most studies, the ratings are used to compute variables such as absolute accuracy, reflecting the absolute match between the judgment and performance, and relative accuracy/resolution, reflecting the discrimination between various levels of performance across items (Schraw, 2009). Specifically, absolute accuracy may be derived using the Hamann coefficient, whereas relative accuracy may be estimated by the gamma correlation. In addition to Likert-type questions, Linderholm et al. (2008) asked students to explain how they went about making their judgments, which revealed the basis of their judgments.

Methodologically, the questions and timing for measuring ME are similar in paper- and e-based reading. It appears that the latter could have better control of the timing for delayed judgment and at specific reading points such as at each node in a hypertext. However, as an increasing percentage of research and learning is carried out in digital environments, the effect of the media on the measurements is worth investigating. Studies that control samples, text content, and design-based characteristics reveal screen inferiority in subjective and objective measures of ME (Ackerman & Goldsmith, 2011). More discussion will be presented in the “Discussion” section.

Simply put, the literature shows that researchers are interested in three types of ME: judgment of the ease/difficulty of the text, judgment of learning/confidence/prediction of performance, and the basis of the judgment. In the first category, learners estimate the ease or difficulty of the text they are going to read or have just read. The second category is concerned with the judgments on how well they understood/learned the content, and their confidence in or predictions regarding the comprehension tests. The last category investigates on what basis they make judgments. Table 5 summarizes the relevant measurement and context. They have been repeatedly used and appeared to be highly recognized by the research community.

Discussion

The present study investigated how metacognitive components (i.e., metacognitive knowledge (MK), skills (MS), and experience (ME)) were assessed in different reading environments with different text formats. In summary, the content analysis reveals that MK was assessed through questionnaires and was mainly examined with static paper-based material. Moreover, ME in reading paper- and e-based static materials was assessed mostly by questionnaires and occasionally by interviews. Unlike MK and ME, MS was assessed by a variety of methods, catching verbal and behavioral data, and was almost equally often investigated in paper- and e-based environments. New indicators have been continuously created to detect students’ use of ME. Tables 3, 4, and 5 illustrated reliable measurement ready for use to assess subconcepts of MK, MS, and ME. The following discussion focuses on Research Question 1 concerning some critical issues of validity of the measurement and Research Question 2 regarding the contribution of measurement in e-reading environment.

When choosing between thinking-aloud and interview/written protocols for assessing MS, researchers should be aware of some constraints regarding their online or offline information generation. Thinking-aloud protocols reflect the information available in respondents’ short-term memory. Although thinking-aloud may place extra cognitive load on learners, it was considered to more accurately depict learners’

utilization of MS. Interview/written protocols, on the other hand, reflect memory traces of the process retrieved from short-term memory or long-term memory after the task is completed (van Gog, Paas, van Merriënboer & Witte, 2005). Thus, some metacognitive skills may have been used but were then omitted due to forgetting, or become fabricated due to memory distortion during interviews. To reduce the influence of forgetting and memory fabricating, records of engagement with that task, such as videotapes, worksheets (e.g., see Butcher & Sumner, 2011, for methods to analyze their essay problems), log files, or students' eye movements (e.g., see van Gog et al., 2005, for making inferences between paragraphs or revisiting paragraphs previously read for monitoring or evaluation) can be used as recall stimuli when assessing MS using these offline measurements.

Moreover, the literature shows that judgments of learning/performance/comprehension induce metacognition, which enhances performance. This suggests an interaction between measurement of ME and performance. Even if the learners are asked to make the judgments after reading, which means that they are no longer able to regulate their effort and time on the text, it still enhances their performance on the test (Palmer, 2003). The judgments seem to activate retrieval processes or in-depth thinking. Researchers have not developed a measurement that could detect students' metacognitive experiences without arousing these experiences.

Furthermore, the advancement of information technology enriches the reading experience of readers. Internet technology affords collaboration, interaction, and problem- or inquiry-based activities to foster meaningful learning. Simulations and animated cues in hypertext have the potential to enhance knowledge construction (e.g., Lin & Atkinson, 2011), whereas integrating game features lifts motivation (Chen, 2014). However, these powerful platforms and features may also drastically increase the complexity and learning time (Gerjets, Scheiter, Opfermann, Hesse & Eysink, 2009) as well as information flow. Thus, dynamic reading in an e-based environment may demand greater mental effort and a higher level of learner control (e.g., Gerjets et al., 2009). Metacognitive knowledge and self-regulation play a decisive role. Our review of research on reading paper- and e-based science texts can shed light on the research design of reading in computerized/digital environments. In the following discussion, we reflect on how assessing metacognition in reading in the computerized/digital environment provides a research context to advance research in SRL.

Digital environments allow researchers to collect a wide variety of data for fixed texts, hypertexts, and Internet searches. First of all, instruments such as questionnaires or survey questions can be integrated with online learning materials for seamless assessment. Moreover, data obtained from the integrated online surveys can be easily aggregated for sequential analysis of data.

Secondly, computer logs and screen recording can make precise MS assessment possible. Although static material could be presented either on paper or on screen, the latter environment provides additional tools for measuring situated metacognitive behaviors. Researchers can pinpoint the type, frequency, quality, and temporal sequence of MS unfolded during science reading. Findings from behavioral analysis can assist in profiling learners' SRL processes and in designing personalized scaffolds for MS training. Moreover, detecting changes in ME across learning stages becomes attainable. Researchers can manipulate the timing of making metacognitive judgments

(e.g., immediate or delayed judgments) in an online inquiry activity (Raes et al., 2012). Thus, researchers may generate more insights into understanding learning effects and designing scaffolding to enhance learning performance.

Thirdly, e-based environments support complicated measurement to grasp the dynamic interactions between task and cognitive conditions. Specifically, the three metacognitive components are closely interwoven and should be deliberated to provide a holistic view of SRL. Our analysis results indicate that only 7% of the reviewed studies took two components into consideration at the same time. Studies which utilize multiple assessments for understanding interactions between metacognitive components are valuable because self-regulation is a chain reaction cycle.

Finally, tracing techniques or protocol analysis can be applied to triangulate the findings. Although 30 of the 55 studies in our review were conducted in e-based reading environments, only four documented computer logs. In these four studies, the researchers applied computer logs to capture temporal data of individuals' actions or behaviors during their reading process for assessing MS. However, computer logs alone are inadequate for understanding strategy use and intention of actions. We suggest combining a protocol analysis with tracing techniques to capture verbal and behavioral data and analyzing both collectively. This approach would increase the accuracy and quality of the evidence and reduce individuals' cognitive load while thinking aloud. Participants' actions or behaviors can be captured in computer logs and later used as a visual stimulus for cuing self-reporting of metacognitive actions in a retrospective interview. A few researchers have begun to explore the potential of recording eye movements as behavioral data and have used these data in cued retrospective interviews (e.g., van Gog et al., 2005). Nevertheless, more research has yet to investigate to provide strong empirical evidence regarding above-mentioned applications.

Conclusion and Future Research

In the present study, we systematically searched Web of Science database and examined the methodology applied in selected articles. Specific attention was given to typical types of assessment for each of the three metacognitive components, MK, ME, and MS, with paper- and e-based static and dynamic science reading material. Rather than composing a comprehensive review about metacognitive measurement in literature, in the result sections, we discussed the strengths and limitations of different metacognitive assessments, for instance, learning journals (Glogger et al., 2012), screen shot analysis (Tu et al., 2008), or thinking aloud (Greene et al., 2008), for grasping event-based metacognitive skills. This allows researchers to choose an assessment with adequate sensitivity according to their study interests and task features. Moreover, we demonstrate how to transfer a measurement from a paper-static to an e-dynamic reading context. Furthermore, we pinpoint some methodological issues for implementation and data interpretation in reading. Finally, we illustrate emerging methods to capture the dynamic behaviors in e-based reading environments. A limitation lies in the relatively small pool of empirical studies investigating metacognition in science reading. Using the current searching practices, our findings and discussion were based on only 47 articles (55 studies). Design and implementation of metacognitive assessments should be adjusted or fine-tuned to accommodate learners' characteristics and task features.

However, based on the limited pool of studies, our findings did not account for SRL in reading tasks of various complexities and metacognition of learners of different age levels. The discussion on the feasibility of the various types of metacognitive assessment could have also been biased for the same reason. Future researchers should replicate our review study once more empirical studies have been published.

According to Winne and Hadwin's (1998) SRL model, learning is a recursive process wherein MK, MS, and ME are highly interactive. Specifically, MK provides a repertoire of MSs, and accurate metacognitive judgment of how one just performed (ME) could provide event-specific guidance for choosing and adjusting particular usage of MSs for better performance. The result of the usefulness of MS updates MK, which provides a better suggestion for future events. Still, this theorized recursive process needs to be examined in the light of evidence. Assessing reading in an e-based environment opens new opportunities to carefully examine the recursive process. Triangulation from multiple assessments and multiple-shot measurements in a cyclic fashion are recommended to fully capture metacognitive behavior in complex digital environments.

Research regarding whether learners transfer the metacognitive knowledge and skills just learned to future events and how many training cycles are needed to develop metacognitive competency to reach a satisfactory level is lacking. To address this research gap, learners' adaptation and improvement of metacognitive competency in a cyclic feedback loop should be traced over multiple tasks to explore the effectiveness of certain instructional designs. In digitalized environment, the manipulation of serial reading tasks is comparatively convenient than in a regular classroom. The results would also shed light on determining better timing for introducing and fading metacognitive scaffolds. For example, Wang (2015) observed that, when introducing evaluation standards as a metacognitive scaffold, the intervention did not have the same effect on calibrating judgment accuracy for different elements of scientific explanation. Underestimation of claims was calibrated in one instructional unit; however, adjusting overestimation of evidence and reasoning required at least five cycles of learning. Similar study designs should be conducted to fill the gap in the field of reading research.

References

Note. Reviewed articles are marked with an asterisk (*)

- Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied*, *17*(1), 18–32. doi:10.1037/a0022086.
- Akturk, A. O., & Sahin, I. (2011). Literature review on metacognition and its measurement. *Procedia- Social and Behavioral Sciences*, *15*, 3731–3736. doi:10.1016/j.sbspro.2011.04.364.
- *Artino, A. R., Cleary, T. J., Dong, T., Hemmer, P. A., & Durning, S. J. (2014). Exploring clinical reasoning in novices: A self-regulated learning microanalytic assessment approach. *Medical Education*, *48*(3), 280–291. doi:10.1111/medu.12303.
- Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning*, *4*(1), 87–95. doi:10.1007/s11409-009-9035-7.
- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, *96*(3), 523–535. doi:10.1037/0022-0663.96.3.523.

- Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist, 45*(4), 210–223. doi:10.1080/00461520.2010.515934.
- Bannert, M., & Mengelkamp, C. (2013). Scaffolding hypermedia learning through metacognitive prompts. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 171–186). New York, NY: Springer.
- Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale, NJ: Lawrence Erlbaum Associates.
- *Burkett, C., & Azevedo, R. (2012). The effect of multimedia discrepancies on metacognitive judgments. *Computers in Human Behavior, 28*(4), 1276–1285. doi: 10.1016/j.chb.2012.02.011.
- *Butcher, K. R., & Sumner, T. (2011). Self-directed learning and the sense making paradox. *Human-Computer Interaction, 26*(1–2), 123–159. doi:10.1080/07370024.2011.556552.
- Chen, Z.-H. (2014). Exploring students' behaviors in a competition-driven educational game. *Computers in Human Behavior, 35*, 68–74. doi:10.1016/j.chb.2014.02.021.
- *Chularut, P., & DeBacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology, 29*(3), 248–263. doi: 10.1016/j.cedpsych.2003.09.001.
- *Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly, 42*(2), 214–257. doi: 10.1598/RRQ.42.2.2.
- Devolder, A., van Braak, J., & Tondeur, J. (2010, August). *Supporting self-regulation in computer-based learning environments: A systematic review of the impact of scaffolding*. Paper presented at the European Conference of Educational Research, Helsinki, Finland.
- *Dhiebi-Henia, N. (2003). Evaluating the effectiveness of metacognitive strategy training for reading research articles in an ESP context. *English for Specific Purposes, 22*(4), 387–417. doi:10.1016/S0889-4906(03)00017-6.
- *DiBenedetto, M. K., & Zimmerman, B. J. (2013). Construct and predictive validity of microanalytic measures of students' self-regulation of science learning. *Learning and Individual Differences, 26*, 30–41. doi:10.1016/j.lindif.2013.04.004.
- Dinmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14. doi:10.1016/j.learninstruc.2012.06.001.
- Donker, A. S., de Boer, H., Kostons, D., Dignath van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review, 11*, 1–26. doi:10.1016/j.edurev.2013.11.002.
- Efkliides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review, 1*(1), 3–14. doi:10.1016/j.edurev.2005.11.001.
- *Franco, G. M., Muis, K. R., Kendeou, P., Ranellucci, J., Sampasivam, L., & Wang, X. H. (2012). Examining the influences of epistemic beliefs and knowledge representations on cognitive processing and conceptual change when learning physics. *Learning and Instruction, 22*(1), 62–77. doi:10.1016/j.learninstruc.2011.06.003.
- *García, F. C., García, A., Berbén, A. B. G., Pichardo, M. C., & Justicia, F. (2014). The effects of question-generation training on metacognitive knowledge, self regulation and learning approaches in Science. *Psicothema, 26*(3), 385–390. doi:10.7334/psicothema2013.252.
- Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior, 25*(2), 360–370. doi:10.1016/j.chb.2008.12.015.
- *Glogger, I., Schwonke, R., Holzäpfel, L., Nückles, M., & Renkl, A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology, 104*(2), 452–468. doi:10.1037/a0026683.
- Greene, J. A., Robertson, J., & Costa, L. J. (2011). Assessing self-regulated learning using think-aloud protocol methods. In B. J. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 313–328). New York, NY: Routledge.
- *Greene, J. A., Costa, L. J., Robertson, J., Pan, Y., & Deekens, V. M. (2010). Exploring relations among college students' prior knowledge, implicit theories of intelligence, and self-regulated learning in a hypermedia environment. *Computers & Education, 55*(3), 1027–1043. doi: 10.1016/j.compedu.2010.04.013.

- *Greene, J. A., Moos, D. C., Azevedo, R., & Winters, F. I. (2008). Exploring differences between gifted and grade-level students' use of self-regulatory learning processes with hypermedia. *Computers & Education*, 50(3), 1069–1083. doi:10.1016/j.compedu.2006.10.004.
- Hsu, Y.-S., Yen, M.-H., Chang, W.-H., Wang, C.-Y., & Chen, S. (2016). Content analysis of 1998–2012 empirical studies in science reading using a self-regulated learning lens. *International Journal of Science and Mathematics Education*, 14(Suppl 1), S1–S27. doi:10.1007/s10763-014-9574-5
- *Ikeda, K., & Kitagami, S. (2012). The effect of working memory capacity and mental effort on monitoring accuracy in text comprehension. *Psychologia*, 55(3), 184–193. doi:10.2117/psysoc.2012.184.
- Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist*, 22(3–4), 255–278. doi:10.1080/00461520.1987.9653052.
- Keck, D., Kammerer, Y., & Starauschek, E. (2015). Reading science texts online: Does source information influence the identification of contradictions within texts? *Computers & Education*, 82, 442–449. doi:10.1016/j.compedu.2014.12.005.
- Koriat, A., & Levy-Sadot, R. (2000). Conscious and unconscious metacognition: A rejoinder. *Consciousness and Cognition*, 9(2), 193–202. doi:10.1006/ccog.2000.0436.
- *Kramarski, B., & Michalsky, T. (2010). Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge. *Learning and Instruction*, 20(5), 434–447. doi:10.1016/j.learninstruc.2009.05.003.
- *Lauterman, T., & Ackerman, R. (2014). Overcoming screen inferiority in learning and calibration. *Computers in Human Behavior*, 35, 455–463. doi:10.1016/j.chb.2014.02.046.
- Lin, L., & Atkinson, R. K. (2011). Using animations and visual cueing to support learning of scientific concepts and processes. *Computers & Education*, 56(3), 650–658. doi:10.1016/j.compedu.2010.10.007.
- *Linderholm, T., Zhao, Q., Theriault, D. J., & Cordell-McNulty, K. (2008). Metacomprehension effects situated within an anchoring and adjustment framework. *Metacognition and Learning*, 3(3), 175–188. doi:10.1007/s11409-008-9025-1.
- Mason, L., Junyent, A. A., & Tornatora, M. C. (2014). Epistemic evaluation and comprehension of web-source information on controversial science-related topics: Effects of a short-term instructional intervention. *Computers & Education*, 76, 143–157. doi:10.1016/j.compedu.2014.03.016.
- *Michalsky, T. (2013). Integrating skills and wills instruction in self-regulated science text reading for secondary students. *International Journal of Science Education*, 35(11), 1846–1873. doi:10.1080/09500693.2013.805890.
- *Michalsky, T., Mevarech, Z. R., & Haibi, L. (2009). Elementary school children reading scientific texts: Effects of metacognitive instruction. *Journal of Educational Research*, 102(5), 363–374. doi:10.3200/JOER.102.5.363-376
- Miles, J. R., & Stine-Morrow, E. A. L. (2004). Adult age differences in self-regulated learning from reading sentences. *Psychology and Aging*, 19(4), 626–636. doi:10.1037/0882-7974.19.4.626.
- *Moos, D. C., & Azevedo, R. (2009). Self-efficacy and prior domain knowledge: To what extent does monitoring mediate their relationship with hypermedia learning? *Metacognition and Learning*, 4, 197–216. doi:10.1007/s11409-009-9045-5.
- Organization for Economic Co-operation and Development (OECD) (2013). *PISA 2015 draft reading literacy framework*. OECD Publishing. Retrieved from <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Reading%20Framework%20.pdf>
- *Otero, J., Caldeira, H., & Gomes, C. J. (2004). The influence of the length of causal chains on question asking and on the comprehensibility of scientific texts. *Contemporary Educational Psychology*, 29(1), 50–62. doi:10.1016/S0361-476X(03)00018-3.
- *Ozuru, Y., Kurby, C. A., & McNamara, D. S. (2012). The effect of metacomprehension judgment task on comprehension monitoring and metacognitive accuracy. *Metacognition and Learning*, 7(2), 113–131. doi:10.1007/s11409-012-9087-y.
- *Palmer, D. H. (2003). Investigating the relationship between refutational text and conceptual change. *Science Education*, 87(5), 663–684. doi:10.1002/sci.1056.
- *Pieschl, S., Stahl, E., & Bromme, R. (2008). Epistemological beliefs and self-regulated learning with hypertext. *Metacognition and Learning*, 3(1), 17–37. doi:10.1007/s11409-007-9008-7
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, California: Academic Press.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Retrieved from ERIC database. (ED338122)
- *Raes, A., Schellens, T., De Weyer, B., & Vanderhoven, E. (2012). Scaffolding information problem solving in web-based collaborative inquiry learning. *Computers & Education*, 59(1), 82–94. doi:10.1016/j.compedu.2011.11.010.

- Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, 26(1), 113–125. doi:10.1023/A:1003044231033.
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4(1), 33–45. doi:10.1007/s11409-008-9031-3.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475. doi:10.1006/ceps.1994.1033.
- *She, H.-C., Cheng, M.-T., Li, T.-W., Wang, C.-Y., Chiu, H.-T., Lee, P.-Z., . . . Chuang, M.-H. (2012). Web-based undergraduate chemistry problem-solving: The interplay of task performance, domain knowledge and web-searching strategies. *Computers & Education*, 59(2), 750–761. doi:10.1016/j.compedu.2012.02.005
- Shell, D. F., Husman, J., Turner, J. E., Cliffel, D. M., Nath, I., & Sweany, N. (2005). The impact of computer supported collaborative learning communities on high school students' knowledge building, strategic learning, and perceptions of the classroom. *Journal of Educational Computing Research*, 33(3), 327–349. doi:10.2190/787L-BCBQ-20FN-FW6C.
- *Tu, Y.-W., Shih, M., & Tsai, C.-C. (2008). Eighth graders' web searching strategies and outcomes: The role of task types, web experiences and epistemological beliefs. *Computers & Education*, 51(3), 1142–1153. doi:10.1016/j.compedu.2007.11.003.
- van Gog, T., Paas, F., van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11(4), 237–244. doi:10.1037/1076-898X.11.4.237.
- Veenman, M. V. J. (2013). Assessing metacognitive skills in computerized learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 157–168). New York, NY: Springer.
- Veenman, M. V. J., & Elshout, J. J. (1999). Changes in the relation between cognitive and metacognitive skills during the acquisition of expertise. *European Journal of Psychology of Education*, 14(4), 509–523. doi:10.1007/BF03172976.
- Veenman, M. V., van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and learning*, 1(1), 3–14. doi:10.1007/s11409-006-6893-0.
- Wang, C.-Y. (2015). Scaffolding middle school students' construction of scientific explanations: Comparing a cognitive versus a metacognitive evaluation approach. *International Journal of Science Education*, 37(2), 237–271. doi:10.1080/09500693.2014.979378
- *Wang, J.-R., & Chen, S.-F. (2014). Exploring mediating effect of metacognitive awareness on comprehension of science texts through structural equation modeling analysis. *Journal of Research in Science Teaching*, 51(2), 175–191. doi: 10.1002/tea.21131.
- *Wang, J.-R., Chen, S.-F., Fang, I., & Chou, C.-T. (2014). Comparison of Taiwanese and Canadian students' metacognitive awareness of science reading, text, and strategies. *International Journal of Science Education*, 36(4), 693–713. doi:10.1080/09500693.2013.826841.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaum Associates.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Orlando, Florida: Academic Press.
- Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review*, 20(4), 429–444. doi:10.1007/s10648-008-9080-9.
- Yee, P. L., Hsieh-Yee, I., Pierce, G. R., Grome, R., & Schantz, L. (2004). Self-evaluative intrusive thoughts impede successful searching on the Internet. *Computers in Human Behavior*, 20(1), 85–101. doi:10.1016/S0747-5632(03)00042-6.
- *Yore, L. D., Craig, M. T., & Maguire, T. O. (1998). Index of science reading awareness: An interactive-constructive model, test verification, and grades 4–8 results. *Journal of Research in Science Teaching*, 35(1), 27–51. doi:10.1002/(SICI)1098-2736(199801)35:1<27::AID-TEA3>3.0.CO;2-P.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, California: Academic Press.
- *Zhang, M. L., & Quintana, C. (2012). Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education*, 58(1), 181–196. doi:10.1016/j.compedu.2011.07.016.
- Zohar, A., & Barzilai, S. (2013). A review of research on metacognition in science education: Current and future directions. *Studies in Science Education*, 49(2), 121–169. doi:10.1080/03057267.2013.847261.