The Effects of Meta-Cognitive Instruction on Students’ Reading Comprehension in Computerized Reading Contexts: A Quantitative Meta-Analysis

Yi-Chin Lan¹, Yu-Ling Lo² and Ying-Shao Hsu³*
¹Department of Early Childhood Education, University of Taipei, Taipei City, Taiwan // ²Department of Special Education, The University of Texas at Austin, USA // ³Graduate Institute of Science Education, National Taiwan Normal University, Taipei City, Taiwan // lollipop0913@gmail.com // jssabrinalo@gmail.com // yshsu@ntnu.edu.tw
³*Corresponding author

ABSTRACT

Comprehension is the essence of reading. Finding appropriate and effective reading strategies to support students’ reading comprehension has always been a critical issue for educators. This article presents findings from a meta-analysis of 17 studies of metacognitive strategy instruction on students’ reading comprehension in computerized reading contexts. Overall, some instances of metacognitive strategy instruction tended to be more effective than others. Additionally, the effects of the instruction seemed to vary according to participants’ characteristics. Drawing upon the findings of this meta-analysis, we propose recommendations for future research and practice.

Keywords

Meta-analysis, Metacognition, Self-regulation, Reading comprehension, Computerized reading contexts

Introduction

A universal agreement among educators is that the ultimate goal of reading is to comprehend text. As the National Institute of Child Health and Human Development (2000) directly pointed out, “Reading comprehension has come to be the essence of reading” (p. 4-1). Without comprehension, reading is reduced to a mechanistic and meaningless skill (Oberholzer, 2005, p. 22).

The acquisition of reading comprehension skills is critical for every student to gain important information from written texts. However, comprehension is a complex skill; in addition, the skills needed to comprehend texts vary by text form, genre, reader capacity, readers’ prior knowledge, and reading goals (The RAND Reading Study Group, 2002). Thus, acquiring and applying reading strategies is thus in turn critical for students.

Comprehension is facilitated when readers use strategies (Rupley et al., 2009). While some findings have shown that good readers might adopt several effective reading comprehension strategies when reading text or during reading tasks, researchers such as Yuill and Joscelyne (1988) argued that less skilled readers especially benefit from instruction. Thus, it is assumed that learning outcomes in reading are related to the quality of the instruction students receive. Various examples of effective reading instruction and their characteristics, challenges, nature, and other significant features have been mentioned and documented. For example, Rupley et al. (2009) found that new material could be bridged with prior knowledge through an explicitly-instructed, detailed process which includes guided practice.

Among these efforts, metacognition has been identified as a significant factor for text comprehension (e.g., Williams & Atkins, 2009). As Harris’s (1990) groundbreaking conclusion suggested, metacognitive abilities seem to be a differentiating factor between good and poor readers. Therefore, Harris (1990) argued, “there would appear to be some value in teaching students to apply metacognitive strategies” (p. 34). In line with this advocacy, many researchers have been devoted to examining the role of metacognitive strategy instruction in reading comprehension (e.g., Cubukcu, 2008; Dabarera, Renanda, & Zhang, 2014). Yet, only a few particularly aimed at examining how metacognitive strategies helped students comprehend digital texts.

In a world in which electronic reading is becoming increasingly common, the reading platform today has shifted from traditional text to hypertext. Prensky (2001) termed students today as “digital natives.” In his words, they are those who have grown up in a world where technology is ubiquitous. Their libraries are on their laptops and other handheld electronic devices; they typically read electronic books rather than printed books. According to
Puntambekar and Stylianou (2005), this kind of text is nonlinear and flexible, thus requiring learners to engage in cognitive monitoring. When reading such texts, they need to “plan what to read next, and closely monitor ongoing learning” (Puntambekar & Stylianou, 2005, p. 454). In other words, these digital natives “process information fundamentally differently from their predecessors” (Prensky, 2001, p. 1). It then becomes clear that the argument “the same methods that worked for the teachers when they were students will work for their students now” is no longer valid (Prensky, 2001, p. 3).

Nevertheless, Srivastava and Gray (2012) pointed out that little consideration has been given to how the aforementioned shift would help or hinder students’ reading comprehension. Therefore, there is an urgent need to conduct such a study to offer in-service teachers who work with their “digital native” students some practical ways of implementing the most appropriate reading comprehension strategies. This article addresses this necessity by presenting findings from a meta-analysis of published refereed quantitative studies that examine the effects of metacognitive strategy instruction on students’ reading comprehension in computerized reading contexts. The findings from this investigation offer a credible source of information about which form of metacognitive strategy instruction would be more effective in terms of helping students comprehend non-traditional texts. This work also offers a rich source of information on the relationships between students’ important individual characteristics (e.g., language status) and the effect of different metacognitive strategy instruction.

This meta-analysis was designed to answer the following four questions:
- What metacognitive strategies in computerized reading contexts have been investigated?
- What is the effect of metacognitive strategy instruction on students’ reading comprehension in terms of grade level, reading ability level, and their language status?
- What is the effect of metacognitive strategy instruction on students’ reading comprehension in terms of genre of instructional content?
- What is the combination of the effect of metacognitive strategy instruction and the type of computerized reading contexts on students’ reading comprehension?

Method

Data collection

According to Petitti (2000), meta-analysis follows several specific steps. First of all, studies of a topic need to be systematically identified, then filtered according to the inclusion criteria.

Article search procedure

The procedure implemented was a three-step, comprehensive search strategy. First, online databases such as the PsycINFO and Google Scholar were scanned for potential studies.

Descriptors entered into the search engines were combinations of reading comprehension AND metacognition, metacognitive, self-regulated, self-monitoring AND computerized/computer-based/computer-assisted, digital, electronic, multimedia/hypermedia, hypertext, on-line literacy, e-book/e-reading, technology/technological, on-line/web-based, interactive learning environment, new literacy, pad, and Internet. As a side note, the reason that we used metacognition and self-regulated/self-monitoring simultaneously was that these two constructs usually connect with each other or are used interchangeably (e.g., Dinsmore, Alexander, & Loughlin, 2008). Also, given that computerized reading contexts might present in a variety of forms, we used as many terms as possible to refer to this type of environment.

The total number of articles found, after combining the three groups of descriptors, was 57.

The second step was an ancestral search. We examined the references in the 57 articles to ensure that no studies related to our topic were left out. Lastly, in addition to the first two steps, we conducted a thorough manual search of five pioneering journals from 1979 to December 2013: Computers and Education, Computers in Human Behavior, Interactive Learning Environments, the Journal of Literacy Research, and Metacognition and Learning. These
journals were specifically examined because of their prominent role in such fields as metacognition, reading/literacy, and digital learning.

**Inclusion and exclusion criteria**

The following inclusion and exclusion criteria were set to guide the article selection:

- Studies were published between 1979 and 2013. In 1979, Flavell published his classic work *Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry*, in which he detailed the notion of this construct. Since then, metacognition has been studied in various content areas (e.g., science education). As such, searching studies from 1979 forward offers a complete analysis of metacognitive strategy instruction provided to students of various backgrounds.

- Studies were published in refereed journals in English.

- Studies involved explicit metacognitive instruction. Clearly defining metacognition has been a challenging task (Veenman, Van Hout-Wolters, & Afflerbach, 2006). Therefore, as Zohar and Barzilai (2013) suggested, “to state clearly which of the existing multiple theoretical perspectives they adopt and to explain the meaning they designate to the construct of metacognition and to the metacognitive sub-components they use in their work” (p. 2) would be helpful for communicating with others about how we define “metacognitive strategy instruction” in this study. We loosely followed Zohar and Barzilai’s (2013) coding scheme, and only used those studies which met any of the nine practices in their framework. Studies that did not involve any form of metacognitive instruction in Zohar and Barzilai (2013) were therefore excluded (e.g., Srivastava & Gray, 2012).

- Studies that did not use any form of electronic text were excluded (e.g., Houtveen & van de Grift, 2007), as were those that did not include any reading comprehension measures (e.g., Greene et al., 2010).

- Studies that adopted an experimental or quasi-experimental design.

- Studies that provided sufficient statistical or quantitative information to allow calculation of the effect size.

After finalizing the inclusion criteria, we then went back to review the titles and abstracts of the potential 57 articles to determine whether they qualified for further analysis. Those studies which included metacognitive strategy instruction aiming to improve the reading comprehension of electronic texts, and outcome measures which included at least one reading comprehension measure were selected from this potential pool. As a result of this data selection process, the final data set included 17 studies reported in 14 articles (see Appendix) that were peer-reviewed, based on original research, took place in a variety of computerized reading contexts, and involved students of diverse backgrounds. This size was considered acceptable since a) there is no specific rule for minimum number of studies, and b) prominent researchers in the meta-analysis area have used less than 20 studies before (e.g., Mol et al., 2008).

**Data analysis**

**Coding procedures**

An extensive coding sheet was adapted from Bryant (2007) for the purpose of the study. Information from the articles was coded in the coding sheet with the following: (a) underlying concept of the study (i.e., purpose of the research, research rationale, research questions), (b) participant description (e.g., age), (c) methodology (e.g., dependent variable(s), (d) intervention condition, (e) control/comparison condition, and (f) findings. Articles were coded by the first two authors using the coding sheet. The information was then summarized into two tables (see Tables 1 & 2).
Reliability

Once the first two authors coded all 17 studies, 4 (24%) were randomly picked to be coded by the other person. Both coders had more than five years of experience teaching and researching in the field of education. Agreement was counted when both coders had the same or similar information for an item on the coding sheet. The mean agreement was 92.15%, ranging from 91.4% to 94%. Through several panel meetings, the disagreements were discussed and resolved for further analysis.

Effect size calculation

Many calculations of effect size (e.g., Cohen’s d, Hedges’ g) can be used to detect the magnitude of effectiveness (Cohen et al., 2011). In this meta-analysis, we used effect sizes to detect the magnitude and strength of the effectiveness of the metacognitive strategies via the computerized reading context.

Cohen’s d is the standardized difference between two populations. For studies where effect sizes were reported, they were recorded directly from the study (e.g., Dreyer & Nel, 2003). For studies where effect sizes were not reported, Cohen’s d was calculated using information provided from the studies. To calculate Cohen’s d, we took means and standard deviations from the studies, subtracted the control group’s mean from the experimental group and divided it by the pooled standard deviation (e.g., Graesser et al., 2007 Study 1). If the mean and the standard deviation were not available, effect sizes were calculated using the F- or t-test scores provided (e.g., Hathorn & Rawson, 2012, Studies 1 & 2). The magnitude of the effectiveness was decided according to Cohen’s (1988) criteria: d = .80 is a large effect, d = .50 is a moderate effect, and d = .20 is considered a small effect.

Results

To lay out the ground for discussion, we first present the study characteristics of the included studies. Then, the types of metacognitive strategy instruction reported in the studies are presented along with the effects. Effects are also presented by participant characteristics and genre. Finally, we analyze the effect of metacognitive instruction and type of computerized reading context on participants’ reading comprehension for the three most and least effective studies, respectively.

Study characteristics

As mentioned, the search yielded 17 studies reported in 14 articles (see Table 1). Of the 14 articles, only two (14%) were published before 2000, while up to 71% were published between 2005 and 2012. These numbers might reveal that interest in examining the effects of such strategies on improving students’ comprehension of computer-based texts has just begun.

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Duration/ frequency</th>
<th>Experimental group condition</th>
<th>Control group condition</th>
<th>Type of computerized reading context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azevedo et al (2007)</td>
<td>82</td>
<td>40 minutes/ once</td>
<td>ERL: Tutor assistance to scaffold self-regulated learning</td>
<td>SRL: No tutor assistance</td>
<td>Hypermedia</td>
</tr>
<tr>
<td>Dalton et al. (2011)</td>
<td>106</td>
<td>24 sessions/ Twice a week</td>
<td>ICON 1: Comprehension strategy ICON 2: Vocabulary ICON3: 1+2</td>
<td>No control group</td>
<td>ICON SDR environment,</td>
</tr>
<tr>
<td>Reference</td>
<td>Duration</td>
<td>Type</td>
<td>Instruction</td>
<td>Aids</td>
<td>Platform</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>------</td>
<td>-------------</td>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td>Gegner et al. (2009) study 1</td>
<td>Approximately 125 minutes/Once</td>
<td>Comprehension aids</td>
<td>No aids</td>
<td>Online computer program written in Flash</td>
<td></td>
</tr>
<tr>
<td>Gegner et al. (2009) study 2</td>
<td>Approximately 125 minutes/Once</td>
<td>Comprehension aids</td>
<td>No aids</td>
<td>Online computer program written in Flash</td>
<td></td>
</tr>
<tr>
<td>Graesser et al. (2007) study 1</td>
<td>50 minutes/Once</td>
<td>SEEK web tutor: critical stance, inquiry, and self-regulated learning</td>
<td>Navigation only</td>
<td>Google™ search engine and websites</td>
<td></td>
</tr>
<tr>
<td>Graesser et al. (2007) study 2</td>
<td>50 minutes/Once</td>
<td>Prior training and SEEK web tutor</td>
<td>Navigation only</td>
<td>Google™ search engine and websites</td>
<td></td>
</tr>
<tr>
<td>Hathorn &amp; Rawson (2012) study 1</td>
<td>30-40 minutes/Once</td>
<td>1. Global monitoring 2. Adjunct inference questions</td>
<td>Text only</td>
<td>Text on computer, DirectRT</td>
<td></td>
</tr>
<tr>
<td>Hathorn &amp; Rawson (2012) study 2</td>
<td>30-40 minutes/Once</td>
<td>1. Global monitoring 2. Specific monitoring</td>
<td>Adjunct fact questions</td>
<td>Text on computer, DirectRT</td>
<td></td>
</tr>
<tr>
<td>Johnson-Glenberg (2005)</td>
<td>2 weeks/4 times a week, 30 minutes per session</td>
<td>Metacognitive strategies: visual &amp; verbal</td>
<td>Anagram</td>
<td>3D-Reader</td>
<td></td>
</tr>
<tr>
<td>Kramarski &amp; Feldman (2000)</td>
<td>2 weeks (8 lessons)/NR</td>
<td>Metacognitive strategy in an internet classroom</td>
<td>Metacognitive strategy in a regular classroom</td>
<td>An internet environment</td>
<td></td>
</tr>
<tr>
<td>MacGregor (1988)</td>
<td>5-6 weeks/2-3 times per week, 20 minutes per session</td>
<td>1. CTS + electronic dictionary 2. CTS + model literal questions 3. 1+2</td>
<td>Hard copies of reading materials</td>
<td>CTS</td>
<td></td>
</tr>
</tbody>
</table>
Note. NR = Not reported; ESL = English as a second language; EFL = English as a foreign language; ELL = English language learners; EO = English only; ERL: externally regulated learning; SRL: self-regulated learning; ICON: improving comprehension online; SDR: scaffolded digital reading; GALT: Glossing Authentic Language Texts; CTS: Computerized-mediated Text System; SAM-LS: social annotation model learning system; ULS: Universal Literacy Environment.

### Design

The majority of studies (n = 14) employed a true experimental design (Cohen et al., 2011), randomly assigning students or classes to treatment or control conditions. Dreyer and Nel (2003) did not assign students randomly; therefore, it is considered to be a quasi-experimental study. Proctor, Dalton, and Grisham (2007) was a single-group design study without a control group, while Johnson-Glenberg’s (2005) was a within-subject design quasi-experimental study.

### Participants

There were a total of 1,210 participants in the studies analyzed, excluding one study that did not report the number of participants (Mendenhall & Johnson, 2010). The majority of the participants (51%, n = 614) were undergraduate students. Participants from secondary schools numbered 412 (34%), and 184 (15%) were from elementary schools.

Participants’ native languages were usually unspecified (n = 10); however, from the descriptions of the studies, they were inferred to be native English speakers. Three studies stated that the participants were native English speakers (Hathorn & Rawson, 2012 Studies 1 & 2; Lomicka, 1998). Participants in Lomicka (1998) were learning French. Two studies included participants who spoke English only as well as bilingual speakers of English and other languages (Dalton et al., 2011; Proctor et al., 2007). Participants in Dreyer and Nel (2003) were speakers of African languages learning English. Participants in Kramarski and Feldman (2000) were learning English as a foreign language.

### Duration and frequency

The instruction was implemented only once in more than half of the studies (n = 10). The duration varied across these studies, with the shortest instruction lasting between 30 and 40 minutes (Hathorn & Rawson, 2012 Studies 1 & 2) while the longest was 125 minutes (Gegner et al., 2009 Study 1).

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Duration/ Frequency</th>
<th>Strategy Supports</th>
<th>Control Group</th>
<th>Learning System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proctor et al. (2007)</td>
<td>30</td>
<td>4 weeks (12 sessions)/3 times per week</td>
<td>Embedded vocabulary and comprehension strategy supports</td>
<td>No control group</td>
<td>Digital approach of ULE</td>
</tr>
<tr>
<td>Puntambekar &amp; Stylianou (2005) study 2</td>
<td>121</td>
<td>45 minutes/Once</td>
<td>Meta-navigation support</td>
<td>No support</td>
<td>CoMPASS</td>
</tr>
</tbody>
</table>
The instruction in seven studies was implemented more than once (Dalton et al., 2011; Dreyer & Nel, 2003; Johnson-Glenberg, 2005; Kramarski & Feldman, 2000; MacGregor, 1988; Mendenhall & Johnson, 2010; Proctor et al., 2007). Notably, although implementing the instruction several times, Dreyer and Nel (2003) and Mendenhall and Johnson (2010) did not report frequency. The remaining five studies reported instruction frequencies ranging from 2-4 times per week. The average duration was 9.5 weeks (range = 2-13 weeks).

Findings

Findings by metacognitive strategy

The metacognitive strategies implemented by the 17 studies could be categorized into three main groups: (1) regulation, (2) strategy cues with think-aloud, and (3) vocabulary and comprehension support. We will address each group in the following.

There were another four studies focusing on the differences between students who read electronic texts and paper texts. Given this dissimilarity, we will report these studies separately (see Table 1).

Regulation as instruction

Metacognition has two fundamental aspects: knowledge of cognition and self-directed thinking (Johnson-Glenberg, 2005, p. 758). The instruction in seven studies (Azevedo et al., 2007; Graesser et al., 2007 Studies 1 & 2; Hathorn & Rawson, 2012 Studies 1 & 2; Johnson-Glenberg, 2005; Puntambekar & Stylianou, 2005 Study 2) involved elements related to the latter aspect.

Azevedo et al. (2007) compared the effects of externally regulated learning (ERL) and self-regulated learning (SRL) for undergraduate students learning science material via hypermedia. Participants in the ERL group received support from tutors, whereas participants in the SRL group did not. This support was for assisting students to achieve self-regulated learning. After reading the text, while both groups gained knowledge and understanding of the text, significant differences between ERL and SRL were found in labeling the human circulatory system and mental models of the system (ES = .55 and .48, respectively).

Similar to the support provided to the experimental group (i.e., the ERL group) in Azevedo et al. (2007), Hathorn and Rawson (2012, Study 1) intended to test if imbedded questions (i.e., global monitoring and adjunct inference) would promote self-monitoring and mental models in reading scientific text via a software program called DirectRT. The results indicated that asking global monitoring questions (i.e., “In the page you have just read, was anything different to what you thought? If so, why?”) maximized learners’ ability to put information on a diagram and reproduce concept maps of the content when compared to the text-only group without monitoring questions (ES = .79 and .73, respectively).

Puntambekar and Stylianou (2005, Study 2) gave participants metanavigation support (italics in original) to help them navigate and learn science content. As they defined it, metanavigation is “the support designed to enable students to reflect upon and monitor their link selection while navigating through a hypertext system” (p. 469). Hence, unlike Azevedo et al. (2007) and Hathorn and Rawson (2012, Studies 1 & 2), their metanavigation prompts included both questions (e.g., “What science concepts will help you solve the challenge?”) and suggestions (e.g., “Think about what other topic descriptions are related to the topic that you are reading”). Puntambekar and Stylianou (2005, Study 2) found that students who received metanavigation support outperformed students who did not in the concept-mapping test (ES = .61 for concept ratio and .70 for connection ratio).

Since the adjunct inference questions did not have an effect on students in the first study (ES = .21 for diagram and .05 for concept map), in their second study, Hathorn and Rawson (2012) sought to find out whether global monitoring questions or specific monitoring questions promoted memory of concepts and facts, and inference, when compared to the adjunct inference questions. Interestingly, they found that the monitoring questions did not have to be specific; participants in the global monitoring group still demonstrated a significant difference when they were asked inference questions compared to the specific monitoring group and the adjunct fact question group (ES = .94).
Unlike the previous four studies, instead of asking students questions, Johnson-Glenberg (2005) prompted students to ask questions and build mental models to monitor their understanding of the text and search for answers to their own questions. Utilizing a within-subject design, students first read texts with anagrams and then read texts with embedded verbal and visual support. The results indicated that students’ comprehension improved significantly ($ES = .45$) when reading texts with embedded support to generate questions and build mental models. Their re-reading behavior was also significantly increased with a moderate effect ($ES = .42$).

The two studies in Graesser et al. (2007) showed very different results. Basically, these two studies were designed to investigate if a web tutor can assist participants to discern the reliability of websites and learn content from websites. The experimental group in Study 1 had access to the SEEK web tutor (Source, Evidence, Explanation, and Knowledge), while the control group did not. The SEEK web tutor was designed to prompt students to think about and justify the reliability and trustworthiness of the websites provided on a mock Google™ page. The design was intended to encourage metacognitive skills of learning and self-regulated study. In the second study, Graesser et al. (2007) added prior training on critical stance to an experimental group to see whether providing such training in addition to the SEEK Tutor would strengthen the use of a critical stance among learners.

Surprisingly, the results of both studies were not very promising. In Study 1, even though the SEEK tutor group gained more knowledge and understanding than the control group, the effect was not significant ($ES = .37$, ns). Instead, participants in the control group outperformed the experimental group ($ES = 1.20$) on the process of studying the websites. The other measures were not affected by the imbedded support. The results of other measures in Study 2 remained the same as in Study 1.

Interestingly, however, the critical stance (where participants in the experimental group mentioned more critical thinking and causal relationships of why a volcano erupted) showed significant difference in favor of the SEEK web tutor group ($ES = .73$ for study 1 and .57 for study 2). Although the connection was not directly examined, Graesser et al. (2007) suggested that expressing critical stance information might have contributed to this difference.

Strategy cues with think-aloud as instruction

Annotations, one kind of strategy cue, appeared to provoke students to reflect more critically upon the primary text (Wolfe, 2008). Lomicka’s (1998) and Yanguas’s (2009) were the only two studies providing participants with strategy cues (i.e., glosses/annotations) with think-aloud as instruction. These two studies had commonalities. First, they both utilized strategy cues providing annotation via software. Additionally, they were designed to examine the differences among different types of glosses. Moreover, the participants in these two studies were English speakers who were learning French and Spanish, respectively.

However, the two studies produced very different results. The statistical analysis in Lomicka (1998) showed no significant difference between the three groups (all glosses, traditional glosses, no glosses) on the percentage of explanation stated by the 12 participants with think-aloud. Yanguas’s study (2009), which compared textual glosses, pictorial glosses, mixed glosses (experimental groups) and no glosses (control group), revealed significant differences in vocabulary recognition and comprehension between the experimental groups and control group. Large effect sizes (range $ES = 1.69-2.78$) indicated that multimedia annotation strongly supported vocabulary recognition and comprehension. Particularly, the mixed gloss group with textual and pictorial glosses outperformed all other groups in comprehension. Lomicka (1998) only collected think-aloud data from participants. If the study had utilized more rigorous vocabulary and comprehension measures, the result would possibly have looked more similar to that of Yanguas (2009).

In addition to the non-significant outcome, the software tracker data in Lomicka (1998) indicated that students in the full gloss group strongly preferred the traditional glosses (definition and translation), even though they could access other kinds of glosses (i.e., definition in French, images, references, questions, pronunciation, and translation in English). Lomicka (1998) argued that the multimedia annotations may assist reading comprehension and metacommments, but the second language (L2) learners tended to only construct text-based comprehension, which led to a preference for traditional glosses.
Vocabulary and comprehension support as instruction

The third group of studies implemented instruction with both vocabulary and comprehension support (Dalton et al., 2011; Gegner et al., 2009 Studies 1 & 2; Proctor et al., 2007). The vocabulary and comprehension support in these four studies were either developing students’ metacognitive awareness or involving students in metacognitive processes.

Dalton et al. (2011) investigated which of the three strategies is the most effective given the Improving Comprehension Online (ICON) scaffolded digital reading (italics in original, SDR) environment: comprehension strategy only, vocabulary strategy only, and mixed strategies of comprehension and vocabulary. In a standardized measure using the Gates-MacGinitie comprehension subtest, there was no significant difference among these groups ($ES = .08$, ns).

Notably, when using researcher-developed measures, the mixed strategy group only outperformed the comprehension strategy group in the narrative comprehension measure ($ES = .65$). Yet, there was still no significant growth in expository comprehension among these three groups.

Still, all three groups showed significant growth in the Gates-MacGinitie vocabulary subtest ($ES = .33$). Additionally, the vocabulary only group and the mixed group both outperformed the comprehension only group in the researcher-designed vocabulary measure ($ES = .58$ and .96, respectively).

In Proctor et al. (2007), 30 fourth grade struggling readers (English only and English language learners) received a 4-week period of instruction embedded with both vocabulary and comprehension strategy support in a digital Universal Literacy Environment (ULE). Similar to Dalton et al. (2011), the result indicated no significant growth on the Gates-MacGinitie Comprehension pre- and post-tests ($ES = .07$). Yet, unlike Dalton et al. (2011), the participants showed no significant growth on the Gates-MacGinitie vocabulary pre- and post-tests either ($ES = .18$). The disaggregated data by English language learners and English only speakers showed similar trends with no significant gains from pre-test to post-test.

Nevertheless, considering that Gates-MacGinitie was a standardized test that might not be sensitive to the subtle change during a 4-week period of instruction, Proctor et al. (2007) chose to use features (i.e., number of clicks) provided by the digital ULE to calculate the correlation. By doing so, although not significant, they found that vocabulary gains were associated with number of clicks on the glossary.

The two studies in Gegner et al. (2009) investigated whether computer-based assistance on vocabulary and comprehension strategies (i.e., narrated animations, glossary terms, and motivational content) improved high school students’ comprehension, perceived difficulty, affect, and motivation of reading scientific articles. Differing from the results of the previous two studies, Gegner et al. (2009) found large effects on reading comprehension ($ES = .79$ for Study 1, $ES = .82$ for Study 2), and the outcomes were significantly different from those of the control groups ($p < .0001$). Furthermore, participants in the experimental groups considered that scientific articles are less difficult ($ES = .96$ for study 1, $ES = .69$ for study 2). These promising results might be due to multi-dimensional supports, such as visual, audio, vocabulary, and background knowledge.

Computerized environment versus hard copy

Lastly, four studies (24%) investigated the difference between reading on paper and reading in a computerized environment (Dreyer & Nel, 2003; Kramarski & Feldman, 2000; MacGregor, 1988; Mendenhall & Johnson, 2010). We consider that these studies investigated different aspects from previous studies as both experimental and control conditions were on the same computer-based interface. Hence, we grouped these studies together and discuss them in an individual section.

Kramarski and Feldman (2000) and MacGregor (1988) demonstrated similar results: the experimental groups did not outperform the control groups in the comprehension measures.
The experimental and the control groups in Kramarski and Feldman (2000) both received the same metacognitive strategy training (i.e., identifying the task, planning, performing and evaluation). In order to examine the contribution of an Internet environment, the experimental group was trained in the environment and the control group was trained in a regular classroom. The results were interesting; the participants in the experimental group did not outperform the control group in reading comprehension, reading strategies, or metacognitive awareness ($ES = -.27, -.48, -1.10$; respectively). They concluded that the Internet environment did not contribute significantly in terms of improving the students’ English reading comprehension, reading strategies, or metacognitive strategies. By giving the control group hard copy reading materials to examine the effects of a computerized-text system, similarly, MacGregor (1988) found no significant difference between the experimental groups and the control group ($ES = .50$ and $.07$, respectively) in any of the comprehension measures.

Interestingly, the experimental group in Kramarski and Feldman (2000) showed higher motivation to read than those in the control group ($ES = .73$). This phenomenon might be due to the novelty of the Internet environment. Likewise, MacGregor (1988) found that the experimental groups made more gains in two vocabulary tests than the control group when they had access to a computerized-mediated text system (CTS) with embedded electronic dictionary, model literal questions, or both ($ES = .88$). It is logical to believe that the experimental groups improved their vocabulary because they could look words up in the electronic dictionary, and benefited from the literal questions.

Mendenhall and Johnson (2010, Study 3) created a social annotation model learning system (SAM-LS) utilizing HyLighter to support students in cultivating critical thinking, writing, and related literacy skills for freshmen. The control group received hard copy reading materials instead of the HyLighter support. Again, the results indicated that there were no significant differences in measures of reading comprehension, critical thinking, or meta-cognition skills.

Since no significant differences between the two groups were found in any of the measures, to explain the unfavorable results, Mendenhall and Johnson (2010) further discussed the weakness of HyLighter that may contribute to the ineffectiveness of the outcome, such as the unstable environment, not fully capturing the intuition of human interaction, and users having to spend some time exploring its functions.

Dreyer and Nel (2003) showed different results from Mendehall and Johnson (2010, Study 3) and the previous two studies. The purpose of Dreyer and Nel’s study (2003) was to offer an academic course within a technology-enhanced environment. They used Varsite to support the instructional delivery in a university in South Africa. Varsite is a learning content management system with which instructors can create, store, manage, and deliver digital learning contents to learners. Learners can access an electronic study guide (with reading strategies) and other resources, perform assessments, and interact with peers via Varsite. In short, the participants in the experimental group received “strategic reading instruction” from this technology-enhanced environment.

The experimental group in Dreyer and Nel (2003) demonstrated a significant increase in English reading, reading pertaining to their profession (i.e., Communication), and TOEFL ($ES = .89, .78, .80$; respectively) compared with participants who did not have access to Varsite. They found that students who received the strategy instruction via the interactive learning system over 13 weeks greatly benefitted in terms of their reading comprehension and standardized tests.

**Findings by participant characteristics**

Studies included in this meta-analysis involved a wide range of participants. To answer the second research question, findings related to participant characteristics across studies of metacognitive instruction utilizing computerized environments are presented to offer a comprehensive understanding of the effects on participants of different grade level, reading ability, and language status.

**Grade level**

Students ranged in grade level from elementary school through university. To clearly present the effects of metacognitive strategy instruction in terms of participants’ grade level, first, we use the first three groups of studies
in the earlier section to discuss this issue. Then, we move on to those studies where computerized and non-computerized reading contexts were both employed.

First, in seven studies in which regulation was used as instruction, the participants were either undergraduate students (Azevedo et al., 2007; Graesser et al., 2007 Studies 1 & 2; Hathorn & Rawson, 2012 Studies 1 & 2) or 6th-7th graders (Johnson-Glenberg, 2005; Puntambekar & Stylianou, 2005, Study 2). Generally, the results were also very similar, revealing that regulation was an effective form of instruction for improving both older and younger students’ reading comprehension. The results of the two studies in Graesser et al. (2007) were an exception. We will address this phenomenon in more detail in the discussion.

Second, Dalton et al. (2011), Gegner et al. (2009, Studies 1 & 2) and Proctor et al. (2007) used both vocabulary and comprehension support. However, Gegner et al. (2009, Studies 1 & 2) aimed at high school students while the other two studies recruited 5th and 4th graders respectively. Generally speaking, the experimental groups in Gegner et al. (2009, Studies 1 & 2) greatly benefitted from the instruction (ES = .79 and .82 for two comprehension measures respectively). In contrast, the elementary students in Dalton et al. (2011) and Proctor et al. (2007) seemed not to improve their reading comprehension based on the instruction provided.

Then, Lomicka (1998) and Yanguas (2009) both used strategy cues with think-aloud as instruction to improve undergraduate students’ reading comprehension. Yet, as presented earlier, the two studies demonstrated very different results. Since the measure in Lomicka (1998) was not that rigorous, it might be problematic to state that the instruction used did not work on older students.

Lastly, students in the four studies which assigned students to either computerized or non-computerized reading contexts involved mixed age groups: elementary (3rd grade), middle school (8th grade) and undergraduate. Since the participants varied in their grade levels, it was challenging to determine the effects of these strategies according to their age. Essentially, only freshmen in Dreyer and Nel’s study (2003) showed significant growth.

Language status

Among seven studies which specifically reported participants’ language status, four (Dalton et al., 2011; Dreyer & Nel, 2003; Kramarski & Feldman, 2000; Proctor et al., 2007) included students who were ESL, EFL, ELL and/or bilinguals. We chose to focus on these studies to highlight the effects of metacognitive strategy instruction on students’ reading comprehension in terms of their language status.

Dalton et al. (2011) and Proctor et al. (2007) included students who spoke only English and those who were English-Spanish bilingual or other bilingual students. Proctor et al. (2007) analyzed students who spoke only English, and English language learners who spoke Spanish, and found that their pre-post comprehension growth after 4 weeks of instruction was not significantly different from each other (ES = .07). Dalton et al. (2011) found a similar trend, that is, there were no significant differences in the Gates-MacGinitie pre-post measures of comprehension between English only speakers and bilingual students who spoke languages other than English.

Notably, in a researcher-created measure on expository comprehension, Dalton et al. (2011) found that English only students outperformed bilingual students who spoke Spanish across conditions (ES = .36, p = .03). The effect might be influenced by English only students’ outstanding performance of vocabulary prior to the instruction.

In Dreyer and Nel (2003) and Kramarski and Feldman (2000), students were either ESL or EFL. As presented earlier, the control groups in both studies did not read in a computer-based environment. However, the findings differed greatly. Dreyer and Nel (2003) concluded that Varsite, a learning content management system, supported students’ learning, and their comprehension measures in English and other areas increased (ES = .78 – .89). At the opposite end, Kramarski and Feldman (2000) concluded that Internet access only helped increase students’ motivation, but not their comprehension, strategy use, or metacognitive awareness (ES = -.27 – -1.10).
Reading abilities

Six studies (35%) specified participants’ reading abilities. Gegner et al. (2009) recruited participants in AP (advanced placement) Biology classes. In contrast, Johnson-Glenberg (2005) recruited poor comprehenders, reading below the mean on the state’s standardized reading test or below the mean for the class in text comprehension. Participants in Proctor et al. (2007) were also struggling readers performing at the 23rd percentile in reading vocabulary and the 31st percentile in reading comprehension on the Gates–MacGinitie Reading Achievement Test.

MacGregor’s (1988) and Dreyer and Nel’s (2003) studies involved students with different reading abilities. Participants in the former were 3rd grade students reading at the 4th to 6th Stanine (average readers) or at the 6th Stanine (good readers). Dreyer and Nel (2003) disaggregated outcomes for successful readers (scoring above 55%) and at-risk readers (scoring below 55%).

A significant effect on the comprehension measure after instruction was observed in Gegner et al. (2009, Studies 1 & 2), indicating that comprehension aids in the form of narrated animations and glossary were effective for students in AP Biology classes (ES = .79 and .82, respectively). Dreyer and Nel (2003) had similar findings for their successful readers after using Varsite as a computer-based tool to deliver studying strategies and content (ES = 1.01–1.14). Yet, in MacGregor (1988), improvement of third grade students with good and average reading abilities was only observed in vocabulary measures, not in comprehension measures.

Essentially, the effects for at-risk or struggling students were less optimal. First of all, although struggling students in Proctor et al. (2007) improved in vocabulary and comprehension, none of them was significant (ES = .06 and .15). While participants in Johnson-Glenberg (2005) benefitted from the intervention of 3D-Reader, there was only a moderate effect (ES = .45). According to their further analysis of the at-risk readers, however, Dreyer and Nel (2003) found that those in the experimental groups outperformed the at-risk readers in the control groups in the English reading and TOEFL posttests (ES = .82 and .72), but not in content reading (Communication Reading) (ES = .30). As a side note, these effect sizes were directly cited from Dreyer and Nel (2003).

Findings by genre

Our third research question aimed at examining the effects of the metacognitive strategy instruction in terms of genre of instructional content because, as Lipson and Cooper (2002) pointed out, the skills needed to comprehend text vary by such factors as text form and genre. Therefore, in the following, we present our findings by different genres.

Except for Kramarski and Feldman (2000) and Mendenhall and Johnson (2010), the remaining studies all provided information about genre/text type. As a side note, the text was all written in English, except that Lomicka (1998) used a French poem and Yanguas (2009) used Spanish text.

Lomicka’s (1998) was the only study using literary text. The results indicated no significant difference between participants who read the poem in the experimental group (full glosses) and the control groups (traditional glosses, no glosses).

Expository texts were adopted by eleven studies. More specifically, all but one (i.e., MacGregor, 1988) employed scientific texts for reading. Most of them related to human biology (e.g., the heart and circulatory system) or natural science (e.g., volcanic eruptions). The effect sizes, although ranging from .07 to 2.72, were mostly moderate to large. Overall, the metacognitive strategy instruction appeared to be effective in helping students comprehend scientific texts. The instruction seemed to also benefit students in other aspects when reading scientific texts. For example, participants in Gegner et al. (2009, Studies 1 & 2) considered that scientific articles were less difficult. On the other hand, MacGregor (1988), who was the only one using non-scientific texts, found that the comprehension outcome did not reach a significant difference (ES = .50). A reasonable explanation was that the number of her participants was very small (n = 12).

In those studies with mixed expository/informational and narrative texts, the effects were also mixed. Dalton et al. (2011) found a moderate effect on narrative text comprehension (ES = .65), but a small and non-significant effect on expository text comprehension. Proctor et al. (2007) found a small and non-significant effect on narrative and
informational texts \((ES = .07)\). Yet, Dreyer and Nel (2003), who specified that the texts were academic, showed promising results with communication and ESL reading comprehension \((ES = .78-.89)\).

**Findings by instruction and computerized environment analysis**

To answer the final research question, we took six studies that had the most and the least effect sizes for the reading comprehension measures, and analyzed the instruction and type of computerized environment between the metacognitive strategy instructions and the control conditions.

Studies in which large effect sizes were observed incorporated glossary/vocabulary hyperlinks with definitions displayed by text or picture. Dreyer and Nel (2003) and Gegner et al. (2009, Studies 1 & 2) both implemented reading strategies and self-monitoring/self-check questions to facilitate students’ reading comprehension, while Yanguas (2009) focused solely on providing definitions in English or pictures of Spanish vocabulary. In general, despite the varying difficulties of the computerized interface, all three studies had large effects on reading comprehension.

The computerized interface of the three studies in which non-significant small effect sizes were observed were more complicated than the three studies described above. For example, Graesser et al. (2007, Study 1) asked participants to rate the reliability and relevance of the websites, and Mendenhall and Johnson (2010, Study 3) requested participants to highlight, take notes, and compare notes with peers via the HyLighter interface. Although these activities seemed to assist metacognitive thinking and monitoring, they did not target reading comprehension directly. On the other hand, Kramaski and Feldman (2000) targeted metacognition strategies and contextual clues that improved reading comprehension, but the control group received the same instruction. The only difference between the experimental and the control groups was that the experimental group had Internet access. Hence, the results only indicated that the use of a computerized interface is not a guarantee of improved comprehension.

**Conclusions**

Helping students to comprehend a variety of texts has always been a key issue for teachers (e.g., Vaughn et al., 2013). Recently, a growing body of research has started to focus on investigating reading comprehension processes while reading computer-based texts (e.g., Ertem, 2010). Many of these studies have especially aimed at examining the effects of the strategies provided to students who read digital texts (e.g., Henry, 2006).

The present meta-analysis chose to specifically examine the effects of metacognitive strategy instruction on students’ reading comprehension in computerized reading contexts. When reviewing the results of the 17 quantitative studies, however, the effects of these different examples of metacognitive strategy instruction remain mixed. Generally, some interesting trends may be drawn from this meta-analysis: 1) metacognitive strategy instruction seems to be more effective in assisting students in comprehending scientific texts, 2) regulation is deemed as a more effective form of instruction, 3) students who received instruction did not necessarily outperform their counterparts, and 4) students’ reading abilities played an interesting role.

Science texts are usually considered as more difficult than non-scientific texts. Hence, educators and researchers have put effort into investigating how to assist students in comprehending science readings such as the content of textbooks (e.g., Smith et al., 2010). Our findings are consistent with several previous studies (e.g., Spence, Yore, & Williams, 1999). In studies where science texts were used (e.g., Gegner et al., 2009, Study 2), larger effect sizes for measures of comprehension were found. Additionally, the metacognitive strategy instruction appeared to benefit students in aspects other than comprehension (e.g., motivation). This might not only reaffirm the impact of metacognitive strategies on helping students comprehend science readings, but also confirm that these strategies remain effective even when the text is presented in digital form.

As we stated in the introduction, showing teachers which form of metacognitive strategy instruction was more effective composed the core purpose of this study. According to our findings, regulation seemed to be more effective than other forms of instruction. Be it asking students questions to prompt them to monitor their reading (e.g., Azevedo et al., 2007) or prompting students to ask questions to monitor their reading (i.e., Johnson-Glenberg, 2005),
students were engaged in self-questioning. More specifically, regulation helped students to be more consciously aware of what they were reading; students who received this form of instruction therefore had better performance in the comprehension measures. Besides, although we found a more consistent result (i.e., regulation appeared to be an effective form of instruction for both younger and older students), little is known about whether the effects of metacognitive strategies are influenced by participants’ age. Therefore, future researchers might consider comparing students across different age groups as they receive metacognitive strategies to draw a more rigorous conclusion.

We also noticed an interesting phenomenon: no discernible differences in reading comprehension measures between the experimental groups were found in several studies (e.g., Dalton et al., 2011). In Graesser et al. (2007, Study 1), the control group even outperformed the experimental group in the process of studying the websites. One possible explanation for this occurrence is that these experimental groups’ cognitive load (Sweller, 1988) might have exceeded the capacity of their working memory. For example, in Graesser et al. (2007), the SEEK tutor was a relatively novel tool for the participants; meanwhile, the SEEK group was asked to conduct multiple actions including on-line ratings and structured note-taking tasks. Thus, Graesser et al. (2007) concluded, “Enhancements are needed in training quality, quantity, and/or both” (p. 103).

Yet, this might highlight a need to re-think the implementation as well as the design of computerized reading contexts. Although studies such as Kim (2013) found that the digital peer contributed to students’ performance in the immediate and delayed posttest text comprehension, there were studies which found totally different results. In the four studies which compared the differences between computerized and non-computerized reading contexts, three of them found that those who received the metacognitive strategy instruction from the computerized reading contexts did not outperform those who did not. In essence, it seems that the computerized reading contexts might not be the determining factor. For example, the HyLighter system in Mendenhall and Johnson (2010) emphasized social interactions among students. In other words, if the quality of students’ social interactions was not good enough, students’ reading comprehension might be affected. Combined, instructors’ roles or other factors such as peer interactions might outweigh the computerized reading contexts per se, just as Clark (2000) reminded us, “educational media do not cause learning, but rather educational methods cause learning” (cited in Gegner et al., 2009, p. 80).

The other intriguing finding was the effect of the instruction in terms of students’ reading abilities. As many studies indicated (e.g., Ertem, 2010), teaching poor or struggling readers strategies would be especially beneficial for them. Hence, it is assumed that less skilled readers would benefit more from instruction than those who already have sophisticated reading skills and/or strategies.

Nevertheless, according to our findings, struggling readers demonstrated mixed yet less optimal results. On the other hand, studies that recruited good or successful readers (e.g., Gegner et al., 2009) showed statistically optimal results as well as having much higher effect sizes. These results, which appear to be opposite to those of previous studies, are not that surprising. As Swanson and De La Paz (1998) and others (e.g., Hopkins & Mackay, 1997) have noted, proficient readers usually execute some metacognitive strategies or already employ other reading strategies. Thus, it might be easier for these skilled readers to quickly adopt or apply these new metacognitive strategies based on their existing reading strategies when they are instructed. In contrast, the task difficulty, the new text structure, and/or the relatively short duration of the instruction (see Proctor et al., 2007) might result in a less optimal or even non-significant result for poor/struggling readers.

The only study that reported at-risk students’ significant improvement (i.e., Dreyer & Nel, 2003) employed both computerized and non-computerized reading contexts. Hence, it is reasonable to assume that the strategies embedded in the computerized reading contexts might have contributed to the significant growth. In other words, with a well-designed computerized reading context, the effects of metacognitive strategy instruction on students’ reading comprehension might be enhanced. Nevertheless, as Ertem (2010) directly claimed, “we know very little about specifically which features of electronic text work best for struggling readers” (p. 142). More studies need to be done to ensure the complex interplay between metacognitive strategy instruction and computerized reading contexts, and how that improves poor/struggling readers’ reading comprehension.
Study limitations and implications

Our main limitation was the quantity of the studies included in this meta-analysis. This added difficulties to generalizing a united conclusion from 17 studies, especially because they varied in terms of the participants’ age, language status, foci of their strategies, etc.

Still, since this meta-analysis demonstrated mixed results in terms of the effects of various metacognitive strategies on comprehending non-traditional texts, this points out a need to conduct more empirical studies to determine what kinds of metacognitive strategies and what types of computerized reading contexts would have greater effects on students’ reading comprehension.

Besides, many of the included studies did not provide sufficient description of the participants. Yet, it appears that the role of students’ backgrounds deserves further attention when probing the effects of metacognitive strategy instruction on their reading comprehension. Thus, we urge researchers to provide a detailed description of participants, especially regarding their reading ability and language status, in future studies. By doing so, we believe that those who are interested in conducting a meta-analysis on related issues could draw a more solid conclusion from sufficient data.

The other implication for future research was inspired by our examination of the fourth group of instruction (i.e., Computerized Environment versus Hard Copy). As aforementioned, those metacognitive strategies did not enable participating students to have better comprehension of digital texts. Still, this tentative finding was made on the basis of very limited studies. More evidence on the effects of the same metacognitive strategy when students use electronic versus paper-based interfaces is needed.

The most important implication for practice was that teachers should know that no one instructional model could be recommended for all. Students’ various backgrounds, the characteristics of diverse content areas, and/or the designs of computerized reading contexts might have an influence on the effects of any specific reading comprehension strategy. As such, teachers at different levels should be more aware of the role of these variables and carefully use our findings as a resource in order to choose the “most effective” metacognitive strategy.

References


Appendix


