Calibration of Comprehension in Adolescence:

From Theory to Online Training

Kleider-Tesler, E., Prior, A. & Katzir, T.

Department of Learning Disabilities and the Edmond J. Safra Brain Research Center for the Study of Learning Disabilities

University of Haifa

In press: Journal of Cognitive Education and Psychology

ABSTRACT

The current study examined the effects of a computerized training program on reading comprehension, confidence ratings, and calibration of comprehension in adolescents with poor and good reading comprehension. Ninety tenth graders participated in the study and completed three training sessions. In each session, participants read two expository texts and answered multiple-choice questions. For each answer they gave, participants also rated their confidence. Participants were assigned to one of three online training conditions that differed in the type of immediate feedback provided after each question: (i) *Feedback on performance; (ii) Feedback on performance and on calibration;* (iii) *Feedback on performance with scaffolding (a cue for correcting wrong answers).* Results demonstrated that scaffolding feedback was the most effective training condition, leading to improved comprehension performance and calibration, especially for poor comprehenders. These findings highlight the necessity of developing theoretical and practice models of online feedback interventions for reading comprehension and self-evaluation abilities.

1. Introduction

In today's global society, success is highly dependent on the ability to independently read large amounts of written text. At school, reading comprehension (RC) is one of the most extensively studied academic skills, and high school students are assumed to have already mastered it. However, there is evidence that not all high school students have mastered the skill of reading comprehension (Biancarosa & Snow, 2004; Scammacca et al., 2007), thus underscoring the importance of identifying effective interventions for this age group.

Despite the academic importance of improving reading comprehension in adolescents, it is unclear whether these abilities can benefit from online training and immediate feedback. Most interventions aiming to enhance comprehension skills, especially for older children, focus on direct instruction of specific strategies (Duke & Pearson, 2002; Rosenshine & Meister, 1994). A different approach to educational intervention stipulates that beyond the skills needed, learning can be improved through dynamic assessment and feedback. Studies indicate that implementation of feedback contributes not only to the student's achievements, but also to the development of meta-cognitive skills (Bolger & Önkal-Atay, 2004; Gery, 1991; Hattie, 2012). Furthermore, feedback which includes a cue on how to correct a wrong answer contributes to the student's achievements, and also to the development of skills such as "learning how to learn", e.g., how to implement the cue they received in the next learning episode (Ackerman & Koriat, 2011; Koriat & Ackerman, 2010). Such types of feedback have been used mostly in the context of classroom learning, in the domains of mathematics and science, among young students (Black, Harrison, Lee, Marshall, & William, 2004). Surprisingly, on-line feedback on academic performance is generally ignored as a tool for RC tasks (Chappuis, Commodore, & Stiggins, 2010).

The current study tested an innovative short online training aimed at enhancing reading comprehension, confidence ratings, and calibration of comprehension in adolescents with a range of initial comprehension abilities.

1.1 Reading comprehension

Reading is a complex process in which readers must use multiple linguistic and cognitive skills to extract meaning from text (Van Gelderen, Schoonen, Stoel, De Glopper, & Hulstijn, 2007; Verhoeven & Van Leeuwe, 2008). In order to comprehend, the reader must read words accurately and efficiently, interpret syntactic information, draw on vocabulary and background knowledge, remember what has been read, and have an understanding of the purpose of reading (Cain, Oakhill, & Bryant, 2000; Gottardo, Stanovich, & Siegel, 1996; Sweet & Snow, 2003).

The most widely used model of RC is the "simple view of reading", which describes two core cognitive processes that influence a reader’s comprehension of a text: decoding and oral language comprehension. Decoding is the ability to extract the form of the word from the written representation, whereas comprehension includes the component processes of retrieving words from lexical memory, determining the intended meaning of individual words, and building meaningful discourse (Gough & Tunmer, 1986; Hoover & Gough, 1990; Tunmer & Chapman, 2012).

Beyond the basic linguistic components that are crucial for reading comprehension, later models introduce additional factors. Thus, The Landscape Model (Van den Broek, Young, Tzeng, & Linderholm, 1999) and the Construction-Integration model (Kintsch, 1988, 1998) suggest that proficient readers engage in complex and dynamic allocation of attention as they proceed through a text. These shifts in attention allow readers to identify meaningful connections within the text and connect them to prior knowledge. Aaron, Joshi, Gooden and Bentum (2008) suggested a componential model of reading (CMR). The model is organized into three domains (a) the cognitive domain (word recognition and comprehension), (b) the psychological domain (i.e., motivation and interest, locus of control, learned helplessness, learning styles), and (c) the ecological domain (i.e., home environment and culture, parental involvement). Thus, current conceptualizations of reading comprehension see it as a complex multi-componential process, including important roles for meta-cognitive processes and self-regulated learning (Ackerman & Leiser, 2014; Glover, 1989; Klassen, 2007).

The complex nature of reading comprehension is also reflected in intervention approaches. To date, most reading comprehension interventions in adolescents have focused on teaching useful strategies. Reading strategies are viewed as cognitive or behavioral actions that skillful readers use to repair and reinforce comprehension while reading (Duke & Pearson, 2002; O'Reilly & McNamara, 2007; Rosenshine & Meister, 1994). Several meta-analyses have highlighted the effectiveness of reading strategies, teacher modeling, and using multiple strategies before, during, and after reading, especially for students with learning disabilities (Edmonds et al., 2009; Fagella-Luby, Schumaker, & Deshler, 2007; Gajria, Jitendra, Sood, & Sacks, 2007).

Most of these interventions have focused mainly on reading comprehension as an outcome measure and did not test possible changes to meta-cognitive monitoring, and specifically calibration, despite accruing evidence regarding their importance for successful comprehension.

1.2 Self-regulation and reading comprehension

Self-regulated learning has been defined as students' cognitive and meta-cognitive strategies used to monitor, control, and regulate their learning (Pintrich, 1988; Pintrich & Garcia, 1994). Monitoring activities include checking, tracking attention, and self-testing for understanding. Such monitoring activities alert the learner to breakdowns in attention or comprehension that can then be “repaired” using control and regulation processes (Weinstein & Mayer, 1986). One way to measure monitoring in reading comprehension is to ask individuals to rate their confidence as to whether they gave correct responses, using a numerical Likert-type scale. This is followed by calculating the difference between the confidence rating and actual performance, which leads to a score called calibration of comprehension (CoC) (Nietfeld, Cao, & Osborne, 2005).

Accuracy in confidence judgment is extremely important as it impacts the meta-cognition of learning (Labuhn, Zimmerman, & Hasselhorn, 2010). Studies have confirmed that effective meta-cognitive regulation depends on learners’ ability to reliably monitor their acquired level of knowledge (Gutierrez & Schraw, 2015; Metcalfe & Finn, 2008; Thiede, Anderson, & Therriault, 2003). This skill is essential for learning, since students need to identify the quality and accuracy of their performance in order to be able to efficiently direct their resources and time-on-task towards areas of difficulty.

High school students are expected to be fluent readers and proficient users of strategies for reading comprehension, such as self-monitoring, summarizing, and self-regulation (Beck, McKeown, Sinatra, & Loxterman, 1991; Duke & Pearson, 2009; Lenz, Ellis, & Scanlon, 1996). Whereas young children are typically optimistic in their self-beliefs and exhibit high confidence, adolescents gain the cognitive capacity to be more self-reflective and aware of their own abilities and difficulties (Stipek, 1993). Hence, adolescents’ decline in self-confidence may be a sign of greater self-awareness and of the improvement of calibration skills (Chiu & Klassen, 2009). Though self-evaluations do become more accurate with age (Schneider, 1998), confidence in performance continues to exceed reality in adolescents (Rogers et al., 2002), leading to positive calibration bias (higher confidence than performance).

Individuals with good comprehension monitor their comprehension and identify and correct difficulties that arise in the text (Westby, 2004). On the other hand, poor comprehenders often are not familiar with different comprehension strategies or how to use them (Meltzer, Katzir, Miller, Reddy, & Roditi, 2004). For students with poor reading skills, ongoing difficulties with decoding, fluency, and comprehension may also hinder the allocation of resources to self-regulation and on-line monitoring of performance. In contrast, students with higher self-monitoring scores show more accuracy in monitoring their comprehension performance (Lin, Moore, & Zabrucky, 2001). Dinsmore and Parkinson (2013) argue that not only do lower achieving students show less accuracy and less skill in learning, they also tend to be less likely to know that their performance is inaccurate or to understand how close they are to the desired learning goal.

Studies have found that poor comprehenders frequently overestimate their abilities. This tendency has been observed in children (Ehrlich, Remond & Tardieu, 1999) as well as in adults (Glover, 1989; Klassen, 2007). For example, among undergraduate students, higher achievers show less overconfidence than lower achievers in reading comprehension tasks (Ackerman, & Leiser, 2014; Silawi, Shalhoub-Awwad & Prior, 2019). Kasperski and Katzir (2013) further found that elementary school students who are good comprehenders have high confidence in their achievement, as well as accurate comprehension calibration scores. In addition, they found significant differences in confidence among three groups of readers (high, average, and low), suggesting a relationship between reading skill and confidence rating.

1.4 Effects of feedback on learning

Research suggests that effective feedback should include elements of both verification and elaboration (Mason & Bruning, 2001). *Verification* is the process in which a student confirms whether an answer is correct or incorrect and it can be accomplished in several ways. The most common way involves simply stating "correct" or "incorrect." *Elaboration* feedback can (a) address the topic, (b) address the response, (c) discuss the particular error, (d) provide worked examples, or (e) give gentle guidance. Elaborated feedback usually addresses the correct answer, may explain why the selected response is wrong, and may indicate what the correct answer should be. There seems to be growing consensus that one type of elaboration, response-specific feedback, appears to enhance student achievements, particularly learning efficiency, more than other types of feedback, such as simple verification or "answer until correct" (Corbett & Anderson, 2001; Mory, 2004; Shute, Hansen, & Almond, 2007).

For example, Narciss and Huth (2004) showed that systematically designed elaborative feedback (additional information relating either to the topic, the task, the errors, or the solution) has positive effects on achievement and motivation in subtraction tasks. Formative feedback can effectively reduce the cognitive load of a learner, which is especially helpful for struggling students (Paas, Renkl, & Sweller, 2003). These students may become cognitively overwhelmed during learning due to high performance demands and thus may benefit from supportive feedback designed to decrease the cognitive load. The freed resources allow the learner to use and activate metacognitive strategies. In fact, previous findings showed how the presentation of explanatory feedback such as verbal explanations of students' choices

(Moreno, 2004; Sweller, van Merrienboer & Paas, 1998), reduces the cognitive load for low-ability students faced with a complex problem-solving task.

However, most studies on the effect of feedback on performance have been conducted on cognitive tasks, or tasks that involve knowledge in a specific domain, such as forecasting the value of stocks or biology content (Bol, Hacker, Walck, & Nunnery, 2012; Bolger & Önkal-Atay, 2004). Only few studies have examined the effect of feedback on reading comprehension. For example, Pridemore and Klien (1991) investigated the effects of student control over feedback in a computer assisted instruction (CAI) program. Results suggested that students who received elaboration feedback (verification information and a short explanation) during instruction performed better than students who received only verification feedback ("Yes, that is correct" or "No, that is incorrect").

These are encouraging findings, which emphasize the need for further research to help clarify the role of elaborative feedback in reading comprehension intervention. In addition, it is of interest if feedback on CoC may also alter metacognitive processes and self-regulation abilities in academic tasks.

1.5 Effects of feedback on self-regulation and calibration

Theoretically, there are strong ties between feedback and meta-cognition (Gery, 1991). Accurate, reliable feedback can provide a source of information that may help learners improve their monitoring and their calibration. Online feedback can not only contribute to students' achievements but may also support the development of skills such as "learning how to learn". For example, studies suggest that students base their metacognitive judgments on their internal feedback from task performance (reading a story), such as feeling of knowing, judgment of learning, and confidence (Ackerman & Koriat, 2011; Koriat & Ackerman, 2010).With regard to external feedback, students are typically very interested in receiving constructive feedback to help them judge the quality of their performance and efficiency efforts, and they find this meaningful and useful for future learning (Berry, 2008; Havnes, Smith, Dysthe, & Ludvigsen, 2012).

When learners become aware of a monitoring error through feedback, they may attempt to resolve the discrepancy between their judgment and their performance. Hattie (2012) proposed that successful learning occurs when this gap is decreased. Moreover, Bolger and Önkal-Atay (2004) investigated the role of feedback on performance and calibration in the learning process. They found that college students showed exaggerated confidence in predicting their success on an experimental task (forecasting the value of stocks), but when they received feedback on their performance their confidence significantly decreased and their calibration became more accurate.

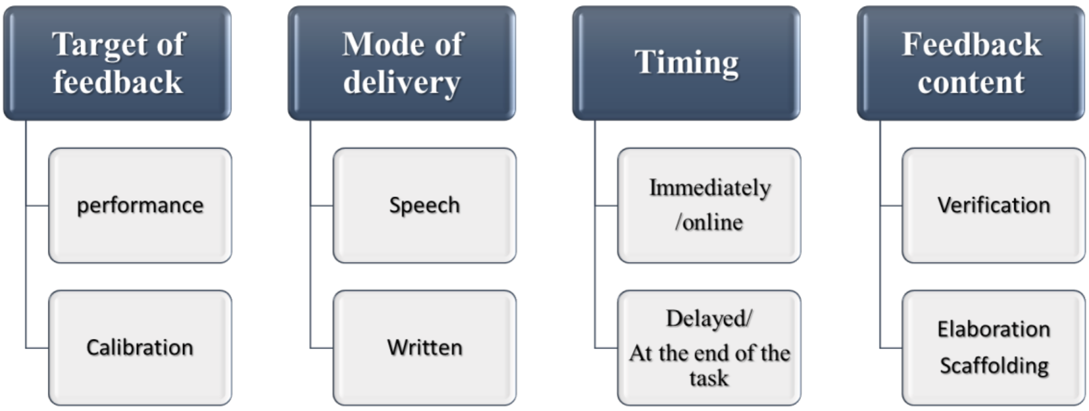
More specifically, Bol, Hacker, Walck and Nunnery (2012) provided high school biology students with monitoring guidelines during a learning phase. The guidelines consisted of questions that targeted how well students were able to: (a) grasp concepts; (b) Assess their confidence in their ability to answer the multiple choice and short-answer items, and (c) Assess their areas of strength, and weaknesses in understanding the test material. Students using the guidelines significantly improved their calibration accuracy and test achievement, thus promoting self-regulated learning. A similar finding suggests that providing corrective feedback may increase students' awareness of the quality of their achievement and thus motivate them to use meta-comprehension processes (Butler & Winne, 1995). Moreover, provision of feedback following incorrect responses was found to be critical for correcting errors, especially when students were over-confident. This finding emphasizes the role of feedback in correcting erroneous performance (Butler, Karpicke & Roediger, 2008).

More evidence has shown that monitoring, calibration, and academic performance (such as multiple-choice tests on course content) indeed improve after feedback on calibration and test performance (Nietfeld, Cao, & Osborne, 2006; O'Connor & Lawrence, 1989; Schunk & Zimmerman, 1998). For example, students who received feedback on their calibration and monitoring accuracy in judging test performance, became more proficient at accurately calibrating their performance on subsequent tests (Nietfeld, Cao, & Osborne, 2006).

An additional type of feedback that has been shown to improve meta-cognitive abilities and to offer students targeted support is scaffolding, which can urge students to engage in attempts to understand challenging material. This scaffolding also encourages students to assume greater responsibility for their learning (Hammond & Gibbons, 2005). There is nearly universal agreement that scaffolding plays an essential and vital role in fostering comprehension (Duffy, 2002; Palincsar, 2003). Different types of verbal scaffolding can include prompts to use particular reasoning strategies (Derry, Hmelo-Silver, Nagarajan, Chernobilsky, & Beitzel, 2006) as well as filling in argument diagrams to learn to distinguish between claims and reasons (Toth, Suthers, & Lesgold, 2002). Feedback that has a scaffolding component has been shown to be more efficient than feedback focused exclusively on performance. An important feature of scaffolding is that it supports students' learning both of how to do the task as well as of why the task should be done that way (Hmelo-Silver, 2006). Indeed, Schwartz and Bransford (1998) showed that providing explanations when needed can be a very effective form of scaffolding (Minstrell & Stimpson, 1996). Scaffolding can also guide instruction and decrease the cognitive load by structuring a task in ways that allow the learner to focus on aspects of the task that are relevant to the learning goals (Hmelo-Silver, 2006; Salomon, Perkins, & Globerson, 1991).

In summary, several studies indicate that different types of feedback may positively affect performance and meta-cognitive abilities at the same time (Bolger & Önkal-Atay, 2004; Nietfeld, Cao, & Osborne, 2006; O'Connor & Lawrence, 1989; Schunk & Zimmerman, 1998). However, recent studies have not focused on complex tasks such as RC, which require multifaceted higher processes. This gap is especially prominent for middle school children, because at this age the requirement for reading comprehension has shifted, especially towards higher order processing which heavily relies on reading strategies. In addition, studies have not examined what specific type of feedback is most effective for improving reading comprehension and calibration. Finally, there is only a limited amount of research on possible differences in the efficacy of formative feedback for good and poor comprehenders. Figure 1 presents a graphic illustration of salient features of beneficial feedback types.

Figure 1: Dimensions of feedback



1.6 Online reading comprehension

Recent decades have seen a sharp increase in the use of computers and electronic devices for reading in both personal and educational settings. Digital media offers users access to seemingly endless sources of information, as well as to reading experiences that are augmented by multimedia and interactive options. Reading in such a setting makes demands on readers' ability to flexibly focus and shift attention and may promote a quick and superficial approach to text reading. Thus, digital media may be less suited for the slower, more time-consuming cognitive, linguistic, and metacognitive processes involved in reading comprehension (Wolf & Barzillai, 2009). Indeed, many students are reading on screen more than in the past, and also report a preference for reading onscreen (Dahan & Brazilai, Katzir, 2018) despite indications of lower performance (Baron, 2017; Delgado, Vargas, Ackerman, & Salmerón ,2018; Mangen et al., 2013).

In addition, presentation modality may also affect readers' self-evaluations of their reading. Recent work has shown that both adults and children are more confident in their comprehension when reading on screen, but exhibit significantly lower comprehension, showing greater gaps between confidence and performance (Ackerman & Lauterman, 2012; Dahan et al., 2018; Mangen, Walgermo & Brønnick, 2013). In sum, reading on screen can be an incredibly useful educational tool, but more research is needed before conclusions can be drawn about its positive and negative features (Baron, 2017).

1.7 Computerized interventions

Many studies have found a positive correlation between the amount of reading practice and reading achievement (Turkeltaub et al., 2003). However, time spent reading without guidance has only a modest influence on reading achievement. The mere allocation of time to independent reading might have little impact upon reading achievement. Similarly, the mere presence of a computer assessment system in the classroom is unlikely to have a significant impact on reading achievement.

Accordingly, in the field of computerized interventions there are mixed results with regard to the effect of online feedback on learning and performance (Angrist & Lavy, 2002; Boozer, Krueger, & Wolkon, 1992; McGowan, 2018; Rouse & Krueger, 2004; Wenglinsky, 1998). For example, one popular instructional computer program designed to improve language and reading skills is known as Fast For Word (FFW). Findings suggest that while the FFW program may improve some aspects of students' language skills, these gains do not appear to translate into a broader measure of language acquisition or into actual reading skills (Rouse & Krueger, 2004). Moreover, Topping, Samuels and Paul (2007) found that effective implementation of such systems involves not only the monitoring of reading practice, but also provides action for guiding the student towards successful comprehension (Topping, Samuels, & Paul, 2007).

1.8 Conclusions and aims of the current study

As part of our attempt to understand in depth the nature of meta-cognitive processes in reading comprehension, the current study examined whether they can be changed by different types of feedback – feedback on performance, feedback on comprehension monitoring accuracy, and scaffolding feedback. Previous studies suggest that feedback is an essential element of learning, especially when it focuses on the task and provides the student with suggestions or hints (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kluger & DeNisi, 1996). Therefore, reading comprehension seems like a perfect candidate area to benefit from computerized training that includes different types of feedback. However, to date, studies have not examined what specific type of feedback is most effective for improving comprehension and calibration in adolescents, and whether this might differ for students who are good or poor comprehenders. Therefore, our online training is a pilot program that examines the effects of different conditions of online feedback (verification vs. elaborated) on reading comprehension from screen.

Thus, the current study addresses the following 3 overarching questions:

1) What is the most effective type of feedback for improving reading comprehension: Feedback on performance, feedback on calibration, or scaffolding feedback?

2) What is the most effective type of feedback for improving confidence and calibration of comprehension: Feedback on performance, feedback on calibration, or scaffolding feedback?

3) Which students benefit the most from computerized reading comprehension training?

It was expected that the most effective feedback on RC and CoC would be (in descending order): scaffolding feedback, feedback on calibration, and feedback on performance. Second, it was expected that RC and CoC would be more significantly improved in poor comprehenders than in good comprehenders.

2. Method

2.1 Participants

A sample of 90 tenth grade students (22 females, mean age 15.30 years) drawn from two high schools in central Israel with a low to medium socioeconomic status, participated in the current study. Almost all (96.5%) of the participants were born in Israel. Five classes were randomly selected and all students whose parents signed consent forms were included in the study. Participants majored in a variety of topics, including: machinery, electronics, software engineering, chemistry, physics, biology, medical sciences, communications, Arabic, dance, design, art, and physical education.

Poor and Good comprehenders. We divided participants into two groups based on a percentile measure of reading comprehension score on an RC test (Dotan & Katzir, 2018; Shany et al., 2006): Poor comprehenders (scored under 60% accurate on the comprehension test) and Good comprehenders (with accuracy above 60%). According to this designation 40 percent, or 35 participants, were considered poor comprehenders (M =42.85, SD=9.60). The remaining 60% of the sample, 55 participants, were considered good comprehenders (M=69.69, SD=9.14).

2.2 Experimental measures

Demographic questionnaire. Non-identifying personal details were collected from all participants, including their age, gender, mother tongue, place of birth, and date of immigration.

2.2.1 *Literacy and reading measures*

Reading skills. The Hebrew adaptation of the Word Reading Efficiency test (TOWRE; Torgesen, Wagner, & Rashotte, 1999; Katzir, Schiff, & Kim, 2012) was used as a measure of word and pseudo-word reading. The total number of words read accurately in 45 seconds was measured (Cronbach's alpha=.95).

Reading comprehension. There are currently no standardized measures of silent reading comprehension in Hebrew. Based on standardized measures of silent reading comprehension in English: Gray Silent Reading Tests (GSRT) (Wiederholt & Blalock, 2000) and Nelson-Denny Reading Test (Nelson, Brown, & Denny, 1960), a

Hebrew measure was developed. In this version we used 7 narrative texts that were adapted to the culture and level of difficulty. The texts covered topics such as "dangerous insects" and "man who gets others to do what he wants" (see Appendix 1).

For each text there were 5 multiple-choice questions that targeted retrieval of information, inference, and integration. RC scores were calculated based on the number of correct responses. The task was first piloted on a sample of 100 children in 11th grade (Cronbach's alpha = .958) (Kulick, Prior, & Kazir, submitted). In the current study the reliability of the task was .671.

2.2.2 *Online experimental training measures*

Reading comprehension. In each session, participants read two texts and then answered five to eight multiple-choice questions. Overall, across conditions, each student read the same number of texts and answered the same number of multiple choice questions. The texts were taken from language textbooks for 10th grade and adapted by the researchers, and were on average 575 words long (range 434-758). All texts were expository, of a descriptive or argumentative structure. Text topics were: influences of alcohol on the brain, dispersion of cellular antennas, contributions of religious faith to physical health, reading among young people, driving under the influence of alcohol, and effects of global warming. Texts were counterbalanced across sessions for different participants. Questions targeted linguistic knowledge, factual understanding, inferential understanding, and logical reasoning, for example: "The first paragraph describes an absurd (unreasonable) idea. What absurd situation is this?" "What are the changes in climate in recent years"? "In line 17, there is a sentence: 'This should be avoided as far as possible'. What does the word 'this' refer to? ". Questions were similar in structure across sessions. The texts were piloted on 80 students and were judged by other researchers, to ascertain appropriate levels of difficulty and performance accuracy (Cronbach's alpha 0.67-0.87).

Self-evaluation measures

Confidence ratings during the task. Reading comprehension confidence scores were measured in accordance with previous research (Kleitman & Stankov, 2001, 2007; Schraw & Roedel, 1994; West & Stanovich, 1997). Following each question, participants were asked: "How confident are you that you responded correctly?" and indicated their answer on a scale ranging from 0 (very unconfident) to 100 (very confident) .

Calibration of Comprehension. This measure consisted of the value of the difference between the confidence rating during the task and performance for each text. First, we calculated the percent of correct responses per text. Then, we calculated the difference between mean confidence ratings per text and performance (Carlvaho & Yuzawa, 2001). Positive scores indicate overconfidence, and negative scores indicate underconfidence.

2.3 Design and procedure

All tests except the reading baseline measures were administered in a group setting, during school hours. Participants received a booklet containing the baseline RC test and a personal information form. After completing all tests, participants were fully debriefed about the study.

2.3.1 Computerized reading comprehension training

Participants completed three 45-60 minute sessions, spaced at 3-4 day intervals. All sessions took place in a quiet computer classroom, with one computer per student. Participants were told that they would be asked to read two texts in each session and to answer between five and eight multiple-choice questions following each text. They were then told that after each question they should indicate "How confident are you that you responded correctly?". When participants answered a question incorrectly, they were informed of their error and given the option of correcting the answer and were once again asked to rate their confidence. All texts and questions were presented in random order.

Participants were assigned to one of three feedback groups:

1. Feedback on performance (PF): Following each text, participants received feedback on their performance: "You answered correctly X percent of the questions" (N=36).

2. Feedback on calibration (CF): Following each text, participants received feedback on their calibration in order to inform them how well their confidence ratings corresponded with their actual responses: "You answered correctly X percent of the questions, and your average confidence ratings were Y" (N=27).

3. Scaffolding feedback (SF): Following each incorrect answer the feedback was "try again" and the participant received a suggestion on how to correct the answer, such as "look at paragraph 2 line 5". After one more try, if s/he answered incorrectly again, the participant was shown the correct answer and moved on to the next question (N=25). Following each text, participants received feedback on their performance: "You answered correctly X percent of the questions".

Thus, in all training groups feedback was given both at the item level and at the text level. As described above, in all feedback groups when participants answered a question incorrectly they were given the opportunity to correct it at the item level.

The training program was built using Open Sesame software and was presented on desktop PCs. Texts were presented as images on the computer screen, and each text was split into two screens (participants could page back and forth among the two text screens, before moving on to read and respond to the questions). Participants could also return to the text from each of the question screens, before selecting their response. Accuracy and calibration feedback were calculated in real time by the computer program and were presented in writing on the computer screen. There was no audio feedback.

A one-way ANOVA was performed and revealed no significant differences between the three training groups on baseline word reading and RC as well as on performance on the first text in the first training session.

3. Results

Although the training study included 3 similar sessions, all the analyses presented below were conducted as pre- and post-intervention analyses, by comparing the first and third sessions across the different conditions.

3.1. What is the most effective type of feedback for improving reading comprehension?

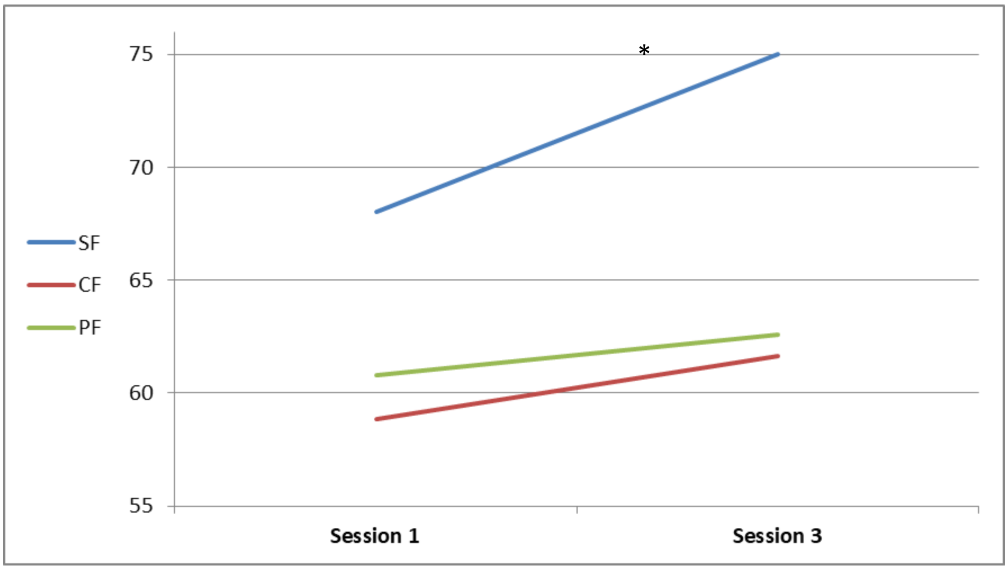
In order to examine this question, a two-way repeated-measures ANOVA was performed, with session (first, third) as the within-participant variable, and feedback group (Performance, Calibration, Scaffolding) as the between-participant variable. The main effect of session was not significant [$F(1,85) = 2.63$, $p > .05$], and neither was the interaction between effects of feedback for the three groups.We ran a planned comparison (with Bonferroni adjusting for multiple comparisons), which revealed that the Scaffolding feedback group showed greater improvement in RC between the first and third sessions, compared to the Performance and Calibration feedback groups ($p<.05$). See Table 1 and Figure 2.

Table 1: Mean comprehension accuracy by sessions and by group

	PF (n=36)		CF (n=27)		SF (n=25)	
	M	SD	M	SD	M	SD
RC session 1	60.8	21.49	58.84	18.18	68	22.06
RC session 3	62.61	18.25	61.64	19.3	75.07	16.36

Feedback groups: PF: Feedback on performance; CF: Feedback on calibration; SF: Scaffolding feedback. Note: The accuracy of comprehension in the first session already includes the training program.

Figure 2: Comprehension accuracy by sessions and by group



Feedback groups: PF: Feedback on performance; CF: Calibration feedback; SF: Feedback with cue after each wrong answer. *p<.05

3.2 What is the most effective type of feedback for improving confidence?

In order to examine this question, a two-way repeated-measures ANOVA was performed on confidence ratings, with session (first, third) as the within-participant variable, and feedback group (Performance, Calibration, Scaffolding) as the between-participant variable. The main effect of session was significant [$F(1,85) = 12.03$, $p < .01$, $\eta^2 = .124$], such that confidence ratings of all three feedback groups increased following training. However, the interaction between feedback group and session was not significant, [$F(1, 85) = .399$, $p > .05$]. Planned comparisons (with Bonferroni corrections) revealed no significant differences between the groups ($p > .05$). See Table 2.

Table 2: Mean confidence ratings by session and by group

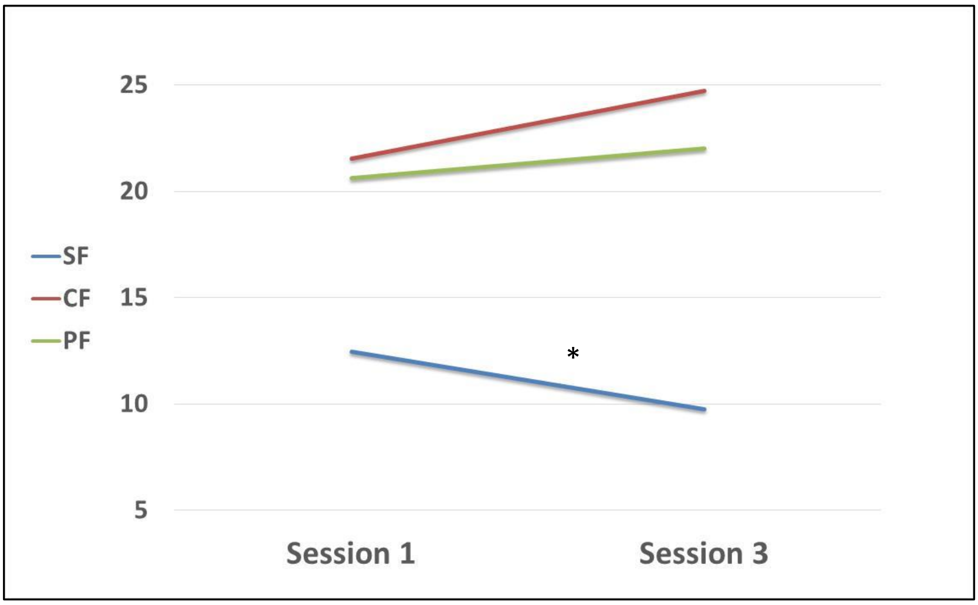
	PF (n=36)		CF (n=27)		SF (n=25)	
	M	SD	M	SD	M	SD
confidence rating session 1	81.42	10.46	80	12.26	80	14.83
confidence rating session 3	84.65	11.44	86	12.84	85	12.42

Feedback groups: PF: Feedback on performance; CF: CoC feedback; SF: Feedback with cue after each wrong answer. Note: The confidence rating in the first session already included the training program.

3.3 What is the most effective type of feedback for improving calibration of comprehension accuracy?

In order to examine this question, a two-way repeated-measures ANOVA was performed on calibration bias, with session (first, third) as the within-participant variable, and feedback group (Performance, Calibration, Scaffolding) as the between-participant variable. The main effect of session was not significant [$F(1,85) = .06$, $p > .05$], and neither was the interaction between feedback group and session, $F(2,85) = .44$, $p > .05$. However, since we predicted different patterns by feedback group, we conducted planned comparisons (with Bonferroni corrections), which revealed that only the group receiving scaffolding feedback showed significant improvement in calibration following training, compared to the performance and calibration feedback groups ($p < .05$). See Figure 3.

Figure 3: Calibration of comprehension by session and by group



Feedback groups: PF: Feedback on performance; CF: CoC feedback; SF: Feedback with cue after each wrong answer.

*p< .05

3.4 Which students benefit the most from computerized reading comprehension training and what is the most effective feedback for them?

First, as described in the participants section we divided the participants into two groups based on a percentile measure of reading comprehension score in the RC screening test (Dotan & Katzir, 2018; Shany et al., 2006).

Then, independent t-tests were performed to examine comprehension, self–confidence, and calibration differences between the two groups in session 1 and 3 (see Table 3). In session 1, poor comprehenders had lower comprehension scores and were less calibrated (more over-confident) than good comprehenders. However, these significant differences disappeared in session 3, such that poor comprehenders improved in RC and CoC. There were no significant group differences in confidence in either session.

Table 3: Comprehension, confidence rating, and calibration bias for good and poor comprehenders, by session

	Poor comprehenders (n=35)		Good comprehenders (n=54)		
	M	SD	M	SD	t(1,87)
RC session 1	54.60	18.15	67.07	20.92	-2.89**
RC session 3	62.14	19.10	68.33	18.19	-1.53
Confidence rating session 1	78.86	11.41	82.02	12.61	-1.19
Confidence rating session 3	83.81	13.29	86.38	11.19	-.98
CoC session 1	24.26	19.88	14.93	18.86	-2.22*
CoC session 3	21.66	19.68	18.03	19.66	-0.84

*p≤.05, ** p≤.01

Then, in order to examine which type of feedback is the most effective for good vs. poor comprehenders, we examined what percentage of participants increased or decreased their comprehension scores between sessions 1 and 3 in each group (see Table 4). Because group sizes were very small when broken down both by feedback type and comprehension level, we include only descriptive statistics here to illustrate the trends, but were unable to analyze the data using traditional statistical tests (such as Chi square).

The pattern of results suggests that in the Performance Feedback group, a larger percentage of poor comprehenders increased their RC than of good comprehenders. In the Calibration Feedback group, the opposite was true – namely, a larger percentage of good comprehenders increased their RC than did poor comprehenders. But the most interesting result was that of the Scaffolding feedback group: 100% of poor comprehenders increased their RC between sessions 1 and 3, whereas only 50% of the good comprehenders showed an increase in comprehension performance.

When comparing the three feedback groups among poor and good comprehenders in terms of the percentage of students who showed a relative increase in RC performance as a result of the training, only among students receiving scaffolding feedback did a larger percentage of poor comprehenders improve their RC compared to the good comprehenders (see Table 4).

Table 4: Percentage of poor and good comprehenders who increased or decreased their RC scores between sessions 1 and 3, in each group

	Poor comprehenders (n=35)		**Good comprehenders (n=53)**	
	Decreased	Increased	Decreased	Increased
PF group	41.2	58.8	58.8	41.2
CF group	62.5	37.5	33.3	66.6
SF group	0	100	50	50

Feedback groups: PF: Feedback on performance (poor comprehenders N=16; good comprehenders N= 20); CF: CoC feedback (poor comprehenders N=12; good comprehenders N= 15); SF: Feedback with cue after each wrong answer (poor comprehenders N=7; good comprehenders N= 18).

4. Discussion

The main finding of the current study is that students receiving scaffolding feedback exhibited the greatest improvement in comprehension and calibration following training. Additionally, all of the poor comprehenders who received this type of feedback showed increased comprehension performance. We shall now discuss these findings in accordance with the research questions.

4.1 *Can a short training improve the RC of adolescents? If so, what type of feedback is the most effective?*

Many different intervention programs target literacy, reading, and cognitive skills among average students and students with difficulties, but a review of these shows that most intervention programs are intended for elementary school aged children, and only a few target high school students (Duke & Pearson, 2002; Fagella-Luby, Schumaker, & Deshler, 2007; Gajria, Jitendra, Sood, & Sacks, 2007; O'Reilly & McNamara, 2007; Rosenshine & Meister, 1994). Moreover, only few intervention programs have used feedback and self-evaluation skills as intervention tools to improve reading comprehension and monitoring abilities.

The current findings indicate that elaborative feedback, i.e., response-specific feedback, appears to enhance student RC, and particularly learning efficiency, more than other types of feedback (see also Corbett & Anderson, 2001; Mory, 2004; Shute, Hansen, & Almond, 2007). Furthermore, feedback that allows for scaffolding learning, by providing readers with a cue on how to correct the wrong answer, contributes not only to the student's achievements, but also to the development of skills such as "learning how to learn", e.g., how to use the cue they received in the next text (Ackerman & Koriat, 2011; Koriat & Ackerman, 2010). This kind of support can promote students' engagement in efforts to understand challenging material. This scaffold also encourages students to assume greater responsibility for their learning (Hammond & Gibbons, 2005) and plays an essential and vital role in fostering comprehension (Berry, 2008; Duffy, 2002; Hammond & Gibbons, 2005; Larson, 2010; Palincsar, 2003; Havnes, Smith, Dysthe, & Ludvigsen, 2012).

All groups exhibited an increase in confidence from the first to the last session. However, only the students in the SF group significantly improved their comprehension. Thus, improvement in the CoC of students who received scaffolding feedback, was driven not just by higher confidence ratings but also by higher comprehension scores.

Surprisingly, the group receiving feedback on calibration did not improve in either comprehension or calibration. These findings are inconsistent with previous studies that have shown that monitoring, calibration, and academic performance improve after feedback on calibration and test performance (Nietfeld, Cao, & Osborne, 2005; O'Connor & Lawrence, 1989; Schunk & Zimmerman, 1998). One possible explanation for this is that participants in the current study did not fully understand the meaning of the calibration feedback. They may not have realized that they can utilize this information in order to attempt to reduce the gap between the confidence rating and accuracy of their performance as much as possible in the next text.

4.2 Patterns of response to intervention by reading ability

The current study adds a novel perspective on the differential impact of training programs on readers with different comprehension levels. As expected, overall poor comprehenders showed lower scores on RC and higher calibration bias (more overconfidence) in session 1, than did good comprehenders. However, these differences were no longer significant in session 3, such that poor comprehenders improved in both comprehension and calibration. Moreover, the results showed that a higher percentage of poor comprehenders who received Performance feedback increased their comprehension than did good comprehenders receiving such feedback. Intriguingly, the opposite pattern was apparent for participants who received feedback on Calibration – in this case, a higher percentage of good comprehenders increased their comprehension than did poor comprehenders. Finally, all poor comprehenders who received scaffolding feedback showed increased comprehension performance, but only half of the good comprehenders improved following this type of feedback.

It seems that poor comprehenders benefit the most from the training, and especially from scaffolding feedback, such that formative feedback can effectively reduce the cognitive load of a learner, especially among struggling students (Paas, Renkl, & Sweller, 2003). These students can become cognitively overwhelmed when learning due to high performance demands, and thus may benefit from supportive feedback designed to decrease the cognitive load. Previous findings have shown how the presentation of worked examples or explanatory feedback (Moreno, 2004; Sweller, van Merrienboer, & Paas, 1998) can reduce the cognitive load for low ability students faced with a complex problem solving task. That is, when students receive an informative cue that directs them to a specific strategy, they can be more engaged and focused on how to answer a question, which allows for meaningful learning.

4.3 Limitations of the study and future directions

The current study was conducted on a small sample size. Future studies should include larger heterogenous samples. In addition, this study focused only on the online training program, and did not include modeling or an explicit teaching component. Future studies should examine whether mini lessons on specific strategies (monitoring, rereading, looking up unknown words) before training add to the depth of learning. Finally, future studies should examine whether the presence of a teacher that is available to support online learning is more beneficial than learning alone without mediation.

Although the training program had a relatively immediate effect on reading, it would also be interesting to examine the long-term impact and how training affects academic performance in the subjects taught at school.

4.3 Conclusions

Results from the pilot training program showed that scaffolding feedback, that is, feedback involving a specific and informative cue, was the most effective. These findings indicate that elaborative, response-specific feedback, appears to enhance student comprehension, and particularly learning efficiency, more than other types of feedback (Corbett & Anderson, 2001; Mory, 2004; Shute, Hansen, & Almond, 2007). Feedback on self–regulation, as delivered in the current study, was not found to be effective, and in some cases also led to a decrease in performance, especially for poor comprehenders. Adolescents may not have understood the nature of the feedback on self-regulation and thus were unable to use it to correct their performance. In the next version of the computerized training program, we are now including an instruction session that defines self-regulation and directly outlines what a student should do in case she receives feedback on overestimation of performance. Finally, the current findings highlight the need to develop theoretical and practice models of RC and meta-cognitive abilities among high school students.

Acknowledgements

This research was supported by grant 1094/14 from the Israeli Science Foundation and by the Edmond J. Safra Brain Research Center for the study of Learning Disabilities. The authors wish to thank Dr. Nachshon Korem for programming assistance, Ms. Sandra Zuckerman for statistics consulting, and two anonymous reviewers for helpful comments on a previous version of the manuscript.

References

Aaron, P. G., Joshi, R. M., Gooden, R., & Bentum, K. E. (2008). Diagnosis and treatment of reading disabilities based on the component model of reading: An alternative to the discrepancy model of LD. *Journal of Learning Disabilities*, *41*(1), 67-84.

Ackerman, R., & Koriat, A. (2011). Response latency as a predictor of the accuracy of children's reports. *Journal of Experimental Psychology: Applied*, *17*(4), 406.

Ackerman, R., & Leiser, D. (2014). The effect of concrete supplements on metacognitive regulation during learning and open-book test taking. *British Journal of Educational Psychology*, *84*(2), 329-348.

Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning. *The Economic Journal, 112*(482), 735-765.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213-238.

Baron, N. S. (2017). Reading in a digital age. *Phi Delta Kappan*, *99*(2), 15-20.

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 26*(3), 251-276.

Berry, R. (2008). *Assessment for learning*. Hong Kong: Hong Kong University Press.

Biancarosa, G., & Snow, C. (2006). *Reading next: A vision for action and research in middle and high school literacy: A report to the Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), -21.

Bol, L., Hacker, D. J., Walck, C. C., & Nunnery, J. A. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology, 37*(4), 280-287.

Bolger, F., & Önkal-Atay, D. (2004). The effects of feedback on judgmental interval predictions. *International Journal of Forecasting, 20,* 29-39.

Boozer, M. A., Krueger, A. B., & Wolkon, S. (1992). Race and school quality since Brown vs. Board of Education. *National Bureau of Economic Research*, *4109*.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245-281.

Cain, K., Oakhill, J., & Bryant, P. (2000). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing: An Interdisciplinary Journal, 13*, 31–56.

Chappuis, S., Commodore, C., & Stiggins, R. (2010). *Assessment balance and quality: An action guide for school leaders* (3rd ed.). Portland, OR: Assessment Training Institute.

Chiu, M. M., & Klassen, R. M. (2009). Calibration of reading self-concept and reading achievement among 15-year-olds: Cultural differences in 34 countries. *Learning and Individual Differences, 19*(3), 372-386.

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 245-252). ACM.

Dahan, D., Brazilai, M., & Katzir, T. (2018). The effects of presentation mode on children's reading preferences, performance, and evaluations. *Computers & Education, 126*, 346-358

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review.*

Derry, S. J., Hmelo-Silver, C. E., Nagarajan, A., Chernobilsky, E., & Beitzel, B. D. (2006). Cognitive transfer revisited: Can we exploit new media to solve old problems on a large scale? *Journal of Educational Computing Research*, *35*(2), 145-162.

Dinsmore, 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.

Dotan, S., & Katzir, T. (2018). Mind the gap: increased inter-letter spacing as a means of improving reading performance. *Journal of experimental child psychology*, *174*, 13-28.

Duffy, G. G. (2002). The case for direct explanation of strategies. *Comprehension instruction: Research-based best practices*, 28-41.

Duke, N. K., & Pearson, D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.). *What research has to say about reading instruction* (pp. 205–242). Newark, DE: International Reading Association.

Duke, N. K., & Pearson, P. D. (2009). Effective practices for developing reading comprehension. *Journal of Education*, *189*(1-2), 107-122.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, *79*(1), 262-300.

Ehrlich, M. F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing: An Interdisciplinary Journal*, *11***,** 29–63.

Fagella-Luby, Schumacher, & Deshler (2007). Embedded learning strategy instruction: Story-structure pedagogy in heterogeneous secondary literature classes. *Learning Disability Quarterly*, *30*, 131-147.

Gajria, M., Jitendra, A. K., Sood, S., & Sacks, G. (2007). Improving comprehension of expository text in students with LD: A research synthesis. *Journal of Learning Disabilities*, *40*(3), 210-225.

Garcia, T., & Pintrich, P. R. (1994). Regulating motivation and cognition in the classroom: The role of self-schemas and self-regulatory strategies. *Self-regulation of Learning and Performance: Issues and Educational Applications*, 127- 153.

Gery, G. (1991). *Electronic performance support systems: Strategic application of technology*. Ziff Institute.

Glover, J. A. (1989). Improving readers' estimates of learning from text: The role of inserted questions. *Literacy Research and Instruction*, *28*(3), 68-75.

Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *Journal of Experimental Child*

Psychology, *63*(3), 563-582.

Gough, P. B. & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.

Gutierrez, A. P., & Schraw, G. (2015). Effects of strategy training and incentives on students’ performance, confidence, and calibration. *The Journal of Experimental Education, 83*(3), 386-404.

Hammond, J., & Gibbons, P. (2005). Putting scaffolding to work: The contribution of scaffolding in articulating ESL education. *Prospect, 20*(1), 6-30.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Oxford, UK: Routledge.

Havnes, A., Smith, K., Dysthe, O., Ludvigsen, K. (2012). Formative assessment and feedback :Making learning visible. *Studies in Educational Evaluation, 38,* 21–27.

Hmelo-Silver, C. E. (2006). Design principles for scaffolding technology-based inquiry. *Collaborative Learning, Reasoning, and Technology*, 147-170.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading & Writing*, *2*, 127-160.

Kasperski, R., & Katzir, T. (2013). Are confidence ratings test- or trait-driven? Individual differences among high, average, and low comprehenders in fourth grade. *Reading Psychology*, *3441*, 59-84.

Katzir, T., Schiff, R., & Kim, Y. S. (2012). The effects of orthographic consistency on reading development: A within and between cross-linguistic study of fluency and accuracy among fourth grade English- and Hebrew-speaking children. *Learning and Individual Differences*, *22*(6), 673-679.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163-182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Klassen, R. M. (2007). Using predictions to learn about the self-efficacy of early adolescents with and without learning disabilities. *Contemporary Educational Psychology*, *32*(2), 173-187.

Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of meta-cognitive processes in test-taking. *Applied Cognitive Psychology*, *15*, 321−341.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences*, *17*(2), 161-173.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254.

Koriat, A. & Ackerman, R. (2010). Metacognition and mindreading: Judgments of learning for Self and Other during self-paced study. *Consciousness and Cognition*, *19*, 251–264.

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, *5*(2), 173-194.

Lenz, B. K., Ellis, E., & Scanlon, D. (1996). Teaching learning strategies to adolescents and adults with learning disabilities. Austin, TX: Pro-Ed.

Lin, L. M., Moore, D., & Zabrucky, K. M. (2001). An assessment of student's calibration of comprehension and calibration of performance using multiple measures. *Reading Psychology*, *22*(2), 111-128.

Mangen, A., Walgermo, B.R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, *58*, 61-68.

Mason, B. J., & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us*. Center for Instructional Innovation, University of Nebraska–Lincoln: 14.

Meltzer, L., Katzir, T., Miller, L., Reddy, R., & Roditi, B. (2004). Academic self-perceptions, effort, and strategy use in students with learning disabilities: Changes over time. *Learning Disabilities Research & Practice*, *19*(2), 99-108.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174-179.

McGowan, C. E. (2018). The most efficient and effective ways to address new literacies for the 21st century classroom. *Theses and Dissertations. 2496.*

Minstrell, J., & Stimpson, V. (1996). A classroom environment for learning: Guiding students' reconstruction of understanding and reasoning. *Innovations in learning: New Environments for Education*, 175-202.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, *32*(1-2), 99-113.

Mory, E. H. (2004). Feedback research review. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Lawrence Erlbaum.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Eds.), *Instructional design for multimedia learning* (pp. 181–195). Munster, NY: Waxmann.

Nelson, M. J., Brown, J. I., & Denny, M. J. (1960). *The Nelson–Denny Reading Test: Vocabulary, comprehension, rate*. Boston, MA: Houghton Mifflin.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education: Learning and Instruction, 74*, 7–28.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance and monitoring accuracy. *Metacognition and Learning, 1,* 159-179.

O’Connor, M., & Lawrence, M. (1989). An examination of the accuracy of judgmental confidence intervals in time series forecasting. *Journal of Forecasting, 8*, 141–155.

O'reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, *43*(2), 121-152.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Palincsar, A. S. (2003). Collaborative approaches to comprehension instruction. In A. S. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 99-114). New York: Guilford.

Pintrich, P. R. (1988). Student learning and college teaching. *New Directions for Teaching and Learning, 33*, 71-86.

Pridemore, D. R., & Klein, J. D. (1991). Control of feedback in computer-assisted instruction. *Educational Technology Research and Development, 39*(4), 27-32.

Rogers, R., Vitacco, M. J., Jackson, R. L., Martin, M., Collins, M., & Sewell, K. W. (2002). Faking psychopathy? An examination of response styles with antisocial youth. *Journal of Personality Assessment, 78*(1), 31-46.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, *64*, 479–530.

Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a “scientifically based” reading program. *Economics of Education Review*, *23*(4), 323-338.

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher*, *20*(3), 2-9.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Schneider, B. H. (1998). Cross-cultural comparison as doorkeeper in research on the social and emotional adjustment of children and adolescents. *Developmental Psychology, 34*, 793-797.

Schraw, G., & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory & Cognition*, *22*(1), 63-69.

Schunk, D. H., & Zimmerman, B. J. (Eds.) (1998). *Self-regulated learning: From teaching to self-reflective practice*. New York: Guilford Press.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475-522.

Shute, V. J., Hansen, E. G., & Almond, R. G. (2007). Evaluating ACED: The impact of feedback and adaptivity on learning. *Frontiers in Artificial Intelligence and Applications*, *158*, 230-237.

Silawi, R., Shalhoub-Awwad, Y. & Prior, A (under review). Comprehension monitoring across L1, L2 and L3: Domain general or language specific?

Stipek, D. (1993). *Motivation to Learn*. Needham Heights, MA: Allyn and Bacon.

Sweet, A. P., & Snow, C. E. (2003). *Rethinking reading comprehension. Solving problems in the teaching of literacy*. New York, NY: Guilford Publications, Inc.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296.

Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66-73.

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (1999). *TOWRE: Test of word reading efficiency*. Austin, TX: Pro-ed.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). “Mapping to know”: The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education*, *86*(2), 264-286.

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, *45*(5), 453-466.

Van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71-98). Mahwah, NJ: Erlbaum.

Van Gelderen, A., Schoonen, R., Stoel, R. D., De Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, *99*(3), 477-491.

Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology, 22*(3), 407-423.

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. Wittrock, *Handbook of Research on Teaching* (pp. 315- 327). New York: Macmillan.

West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin & Review*, *4*(3), 387-392.

Wenglinsky, H. (1998). Does it compute? The relationship between educational technology and student achievement in mathematics. Princeton, NJ: Educational Testing Service.

Westby, C. (2004). A language perspective on executive functioning, meta-cognition, and self-regulation in reading. In C. A. Stone, E. R. Silliman, B. J. Ehren, & K. Apel (Eds.). *Handbook of language and literacy* (pp. 398-427). New York: The Guilford Press.

Wiederholt, J., & Blalock, G. (2000). *Gray silent reading tests*. Austin, TX: Pro-ed.

Wiemer-Hastings, P., & Graesser, A. C. (2000). Select-a-Kibitzer: A computer tool that gives meaningful feedback on student compositions. *Interactive Learning Environments, 8*(2), 149-169.