**Contribution of Metacognitive Questions   
to Accuracy of Judgment in Learning in a Digital Environment**

1. **Introduction**
   1. **Developing Literacy in a Digital Environment**

According the United Nations Educational, Scientific and Cultural Organization (UNESCO), literacy is the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society (UNESCO, 2004).

Development of literacy among students has become a cardinal objective of education systems worldwide. Accordingly, these systems engage in developing curricular materials for furthering literacy skills in various disciplines, including reading literacy and mathematical literacy. Mathematical literacy is an individual’s capacity to reason mathematically and to formulate, employ, and interpret mathematics to solve problems in a variety of real-world contexts. It includes concepts, procedures, facts and tools to describe, explain and predict phenomena. It helps individuals know the role that mathematics plays in the world and make the well-founded judgments and decisions needed by constructive, engaged, and reflective twenty-first-century citizens (OECD, 2018).

Today, technological changes have created digital learning environments that facilitate teaching and learning processes that are substantially different from those invoked by the printed book. These have brought about a shift in reading and learning habits, from those typical of on-paper performance to others characteristic of reading in a digital environment (Murphy et al., 2003; Winter 2018). Learning in a digital environment has many advantages. It is accessible, convenient (anywhere and anytime), and inexpensive; it offers use of diverse media (digital tasks, virtual online learning systems, MOOC courses) and broad fields of pooled content (digital books and journals, shared encyclopedias, and information databases in a range of disciplines), to name only a few (Csapó et al., 2012; Dennis et al., 2016). The research literature, however, has identified learning difficulties that surface in digital environments relative to on-paper formats, which manifest in learners’ achievements in reading comprehension (Daniel & Woody, 2013; Kong et al., 2018; Mangen et al., 2013; Murphy et al., 2003; Singer & Alexander, 2017; Sperling et al., 2002). In an attempt to explain learning inferiority in a digital environment, we tested several hypotheses. Some are associated with technological failures that have a physical effect, such as visual difficulties traced to screen display characteristics (Benedetto et al., 2013; Chou, 2016; Jeong, 2012), as well as a cognitive impact, e.g., loss of mnemonic aids when learners navigate linear digital documents by scrolling (Chou, 2016; Mangen et al., 2013). However, even in studies examining digital reading that avoids these technological failures (e.g., e-books and tablets) and that uses short texts, eliminating the need for scrolling, inferior achievements relative to on-paper learning were observed (Daniel & Woody, 2013; Delgado et al., 2018; Kong et al., 2018; Sidi et al., 2017).

To explain on-screen learning inferiority, researchers claim that readers consider the study of informative contents in digital environments easier because digital environments are commonly used in rapid, superficial, and decentralized interactions (correspondence, posting, etc.), which do not entail cognitive effort and deep processing (Morineau et al., 2005; Mueller & Oppenheimer, 2014). Accordingly, readers tend to develop sloppy and shallow on-screen reading strategies (Liu, 2005). Readers who use digital media intensively struggle to cope with complex informative texts (Annisette & Lafreniere, 2017) and the more frequently they use such media, the more difficulty they encounter (Duncan et al., 2015; Pfost et al., 2013). To determine how reading processes in digital environments may be improved, researchers have been testing metacognitive processes that regulate and monitor learning. One of these processes is judgment.

**1.2 Judgment**

According to the model put forward by Nelson and Narens (1990), individuals self-judge the quality and quantity of the knowledge that they acquire as they acquire it. Judgment is a metacognitive process that people carry out repeatedly, and they monitor and regulate their learning based on this (Dunlosky & Rawson, 2012; Pieger et al., 2016).

According to the model of “narrowing the gap” (Butler & Winne, 1995; Dunlosky & Thiede, 1998; Nelson & Narens, 1990), learning is halted when the learner judges that they have attained an adequate level of knowledge. Additional time is devoted to learning when learners still consider their level of knowledge to be below the target level. It follows that when judgment is biased toward overconfidence, learning may be halted prematurely, whereas under-confidence may result in an overallocation of learning time (Bjork et al., 2013). If so, inaccurate judgment may result in ineffective learning and poor achievements. Accuracy in judgment of comprehension is calculated by comparing judgment with achievement in two senses: relative (resolution) and absolute (calibration). Relative accuracy denotes the extent to which individuals’ judgments correspond to their test achievements (Dunlosky & Thiede, 2000; Thiede, Anderson, & Therriault, 2003) and may be calculated by means of Goodman and Kruskal's gamma. The gamma correlation, ranging from 1 to -1 and ideally around 1, expresses a learner’s ability to differentiate between texts learned well and those not. Absolute accuracy (calibrated accuracy or calibration) is the extent to which judged achievement reflects actual achievement (Bol et al., 2005); it expresses the distance between judged achievement and actual achievement and its direction (above or below actual achievement). The smaller the disparity, the closer calibration is to the ideal.

Ackerman and Goldsmith (2011), comparing on-paper learners with digital-environment learners, found that those among the latter who struggled to self-judge their learning committed misjudgments that manifested in overconfidence in self-assessed knowledge. Namely, their self-assessed knowledge surpassed the knowledge they had actually acquired as reflected in performance on tests. The disparity that came about between judgment and achievement was significantly wider than that between judgment and achievement in on-paper learning. This finding recurred in additional studies among different age groups, learning a variety of subjects, and under different learning conditions (Ackerman & Goldsmith, 2011; Ackerman & Lauterman, 2012; Dahan Golan et al., 2018; Lauterman & Ackerman, 2014; Sidi et al., 2016, 2017). These findings aside, young learners are found to have particular difficulty in accurately judging their own knowledge and correctly assessing how well they understand what they have learned (Pieger et al., 2016; Thiede et al., 2003).

If so, learning in a digital environment is typified by faulty metacognitive monitoring based on inaccurate judgment and is inferior to online learning for this reason. It is found, however, that learners who invoke learning strategies that induce deep processing (such as summarizing paragraphs, writing keywords, etc.) improve their judgment accuracy and achievements and eliminate the digital learning inferiority (Lauterman & Ackerman, 2014; Sidi et al., 2017). Many researchers call for continued testing of innovative pedagogies based on additional learning strategies that may trigger deep processing and improve learning in a digital environment (Delgado et al., 2018; Kong et al., 2 018).

**1.3 Metacognitive Teaching**

Recent years have seen the development of learning strategies based on metacognitive processes. Resting on a broad basis of professional literature, they set within a theoretical and empirical framework the importance of metacognitive teaching in promoting learning processes in traditional and digital environments. It has been found in research that the inculcation of metacognitive processes in learners advances their cognitive abilities in various fields, such as problem-solving and inquiry-based learning (Kim, 2016; Mihalca et al., 2017), reading comprehension (Carretti et al., 2014; Muijselaar et al., 2017), processing information and media (Pellas, 2014; Winne & Hadwin, 2013), and foreign-language studies (Forbes & Fisher, 2015; Pishghadam & Khajavy, 2013). Such teaching has been found to be effective in promoting metacognitive processes among learners and improving their achievements in diverse fields of knowledge and learning environments (Boström & Lassen, 2006; de Boer et al., 2018; Ellis et al., 2014; Kramarski & Kohen, 2017; Michalsky, 2013; Michalsky et al., 2009). One method of metacognitive teaching is the presentation of questions that bring metacognitive processes to the learner’s awareness and are aimed at monitoring and regulating learning. Directly asking metacognitive questions significantly improves learners’ achievements in text comprehension (McKeown et al., 2009; McMaster et al., 2010; Michalsky et al., 2009). The question that emerges is: Does metacognitive self-questioning improve judgment of learning accuracy in a digital environment? This question is amplified by to the dearth of studies that examine the various aspects of accuracy in judging comprehension among young learners, despite the fact that young learners are capable of regulating their learning by applying monitoring and judgment (Baars et al., 2014). Accordingly, our goal in the present study is to investigate the effect of asking metacognitive questions on young learners’ judgment accuracy as they learn texts related to mathematical literacy and reading literacy. We hypothesize that metacognitive support in the form of self-questioning will improve judgment accuracy, mitigate on-screen learning inferiority, and enhance learners’ achievements in reading and mathematical literacy in a digital environment.

1. **General Description of the Two Studies**

The research described below comprises two studies that tested the effect of support, given in the form of asking metacognitive questions about the accuracy of judgment regarding comprehension in digital learning in two fields of knowledge: reading literacy and mathematical literacy.

In both studies, learners read texts on screen, were asked to judge their understanding, and took a reading comprehension test. An intervention program was developed in order to examine the effect of metacognitive support on judgment accuracy. This program integrated metacognitive self-questioning into the environment in which the text was read.

The purpose of Study 1 was to test the effect of metacognitive support on judgment accuracy in digital learning for reading literacy. This study used lengthy texts, which primary-school students were asked to read.

The purpose of Study 2 was to test the effect of metacognitive support on judgment accuracy in digital learning for mathematical literacy. This study, also with primary-school students, used short texts that included mathematical language and symbols.

1. **Study 1**

**3.1 *Method***

**3.1.1 Participants**

Sixty-five fifth-graders (ages 10–11, 52.3% girls) took part in this study. They attended two different schools that were invited to take part in the study and gave their consent. The schools were of similar socioeconomic status as defined by the Israel Central Bureau of Statistics (2019). Students who defined themselves as recent immigrants were omitted from the sample because their reading comprehension difficulties might be traced to the challenges of reading in a foreign language.

The students were divided randomly into two groups: Group 1, “Meta\_read” (*n*=32, 62.5% girls) engaged in lessons on reading literacy, and received metacognitive support in the form of questions integrated into the text. Group 2, “Control\_read” (*n*=33, 42.4% girls), engaged in lessons on reading literacy without receiving support.

**3.1.2 The Intervention Program**

To test our hypotheses, a courseware product called “*Oriani—ze ani!*” (Literacy—That’s Me!) was developed. The program presents learners with six different texts that test their reading literacy capabilities, tailored to the fifth-grade curriculum. The texts were taken from tests by the National Authority for Measurement and Evaluation in Education (RAMA) in Israel. To provide metacognitive support, four metacognitive self-questions following the IMPROVE model (Mevarech & Kramarski, 1997) were integrated into the texts: a comprehension question (“What is discussed in the text that I read?”), a context question (“What topic that I am familiar with is related to this text?”), a strategy question (“What strategies will I use to help me understand the text that I read?”), and a reflection question (“I’ve finished reading. Did I understand the text? How can I improve my reading next time?”)

**3.1.3 Research Tools**

A courseware product called “*Oriani—ze ani!*” (Literacy—That’s Me!”) was developed for the purposes of this study, written with the C Sharp programming language. It enabled students to read texts, judge their comprehension, and be tested.

Figures 1–4 present the texts, the judgment question, and the test questions.

Insert Figures 1-4 here

*Measuring Accuracy in Judgment of Comprehension*

After reading each text, the participants were asked to judge how well they understood it. There are several possible ways of posing this type of question (Ackerman & Goldsmith, 2011; Pilegard & Meyer, 2015; Redford et al., 2012). In the current study, participants were asked a performance-prediction question: “You are about to answer seven questions related to the text. How many of the seven will you be able to answer correctly?” The answer was given by sliding a cursor along a ruler to measure their comprehension on scale of 0–7.

After judging their own comprehension, the courseware presented participants with a multiple-choice test that examined their comprehension of the text. As stated, judgment of accuracy was calculated by comparing their assessed comprehension with test scores in two respects: resolution and calibration. In this study, six judgments were compared with outcomes on six tests.

*Evaluating Achievements in Reading Literacy*

Achievements in reading literacy were evaluated by having the participants read and study six literacy texts in Hebrew and take reading-comprehension tests. There were various types of texts (fictional, scientific, etc.), and their average length was 650 words. To evaluate comprehension, a seven-item multiple-choice test was tailored for each text. The texts and their corresponding tests were based on the “Meitzav” (Hebrew acronym for “School Growth and Efficiency Measures”) standardized tests in Hebrew and were adjusted to assure that they were appropriate for the study. The Hebrew-language Meitzav exams that test reading literacy among fifth-graders were developed by RAMA and passed quality testing that established their validity (RAMA, 2009).

A reliability check for internal consistency (Cronbach’s Alpha), performed for the full set of questions on all the tests (42 items), yielded an α=0.840 reliability coefficient. On the basis of this coefficient, a general index score on the comprehension test was calculated for each participant by averaging all 42 items in total. A separate score for each test was also calculated. The scores were on a scale of 0–7, with higher scores indicating a higher level of achievement.

**3.1.4 Procedure**

Participants were taught how to use the courseware in the computer room. The participants in the “Meta\_read” group also received an explanation about metacognition in learning and practiced the use of metacognitive questions as they read texts involving reading literacy in a digital environment.

Afterwards, all participants attended a workshop comprised of weekly one-hour lessons in which they were asked to work by means of the courseware. Participants in the “Meta\_read” group read a text in which the following metacognitive questions were inserted: “What is discussed in the text that I read?”; “What topic that I am familiar with is related to this text?”; “What strategies will I use to help me understand the text that I read?”; and, “I’ve finished reading. Did I understand the text? How can I improve my reading next time?” They were asked to answer these questions in writing, using a notebook that they had been given.

After each participant finished reading, he or she pressed an on-screen button to confirm that they completed the task. At this stage, the text disappeared from the screen and was replaced by a performance-prediction question: “You are about to answer seven questions related to the text. How many of the seven will you be able to answer correctly?” The participant answered by sliding a cursor along a ruler with a scale from 0 to 7. Then the following question appeared: “Would you like to read the text again?” If the participant pressed “Yes,” the text reappeared; otherwise, the participant advanced to the testing stage. At this stage, one question and several possible answers appeared on the screen. The participant was asked to select the right answer.

Table 1 summarizes the actions taken at each stage of the study.

To ensure that participants indeed based their judgments on their perceived level of comprehension of the text and not randomly, they were asked to read and be tested on a seventh test that was significantly harder than the others, and not suited to their age. This was meant to rule out the possibility that participants judged themselves randomly or uniformly across the various texts. If participants indicated a low score to judge their own comprehension of the seventh test, this would indicate a sensible judgment that reflected the level of difficulty of the text.

**3.2 Results**

**Literacy Achievements**

Table 2 presents the reading literacy achievements of each research group separately.

The participants’ pre-test did not include metacognitive support. To rule out the possibility of variance in the achievements of the research groups, an independent-samples t-test was performed. Indeed, no difference was found in the achievements of the “Control\_read” and the “Meta\_read” groups: *t*(51.5)=1.52, *p*=.133.

To check for significant differences between achievements in the pre-test and those in the post-test, within each group, a two-way repeated-measures analysis of variance was done (points in time: pre- and post-) x 2 (research groups: “Meta\_read” and “Control\_read”). In this analysis, no principal significant difference was found in achievements over time: *F*(1,59)=1.430, *p*=.237, =.024. However, a significant time x group interaction effect was found: *F*(1,59)=4.931, *p*<.005, =.077.

Figure 1 presents the interaction effect graphically:

Insert Figure 1 here

Observation of the averages presented in Figure 1 and the simple-effect analysis reveals a significant difference between the tests among participants in the “Meta\_read,” such that achievements on the post-test surpassed those on the pre-test: *F*(1,30)=7.381, p<.05, = .197. In the “Control\_read” group, no significant achievement differences between the tests were detected.

To check for a significant effect on literacy achievements as manifested in all six tests, an independent-samples t-test was performed, with the average of the six exam scores as the dependent variable and the research group as the independent variable. The results show a significant difference between the research groups in average scores: t(63)=2.967, p<.005, as the achievements of the “Meta\_read” group (*M*=4.169, *SD*=1.12) significantly surpassed those of the “Control\_read” group (*M*=3.26, *SD*=1.34).

**Judgment of Comprehension**

Figure 2 presents the judgments and achievements of the six tests by research groups.

Insert Figure 2 here

To ensure that the judgment was based on perceived levels of text comprehension and not on random answers, the participants were asked to read and be tested on a seventh, more difficult text not suited to their age.

Figure 3 shows the differences between average achievements and judgments on the six tests for the age-suitable texts and those on the seventh text, more difficult text.

Insert Figure 3 here

The data in Figure X [3?] and the paired-sample t-test analyses reveal, as expected, a significant difference between average test achievements on the six age-compatible texts and those on the seventh text: *t*(59)=7.836, *p*<.001, such that the average achievements on the six tests (*M*=3.744, *SD*=1.33) exceeded those on the seventh (*M*=2.55, *SD*=1.443).

Another significant difference was found between the average judgment on the six age-appropriate tests and that on the seventh, *t*(59)=7.393, *p*<.001, such that average achievements on the six tests (*M*=5.497, *SD*=1.264) surpassed those on the seventh (*M*=5.497, *SD*=1.264).

To check for a difference between achievements and judgments on each of the six tests and those on the seventh, paired-sample t-tests were performed. All of them came out significant, such that significant differences were found between achievements and judgments on each of the six tests compared to those on the seventh.

**Absolute Accuracy/Calibration**

One aspect of accurate judgment is calibration, defined as the absolute difference between pre-test judgment and actual test achievements. Table 3 presents the participants’ calibrations on the six tests, divided according to the research groups.

Insert Table 3 here

To examine the relationship between calibration and reading-literacy achievements, the correlation between the absolute disparity and achievement was measured. Here a significant negative relation was found between the absolute gap and the achievements (*r*=-.72, *p*<0.001), i.e., the smaller the disparity—the better the calibration—the higher the achievements were.

The participants’ pre-test did not include metacognitive support. To make sure there was no variance in calibration between the research groups, an independent-samples t-test was performed. Indeed, the “Control\_read” and the “Meta\_read” groups showed no difference in calibration during the pre-test: (50.4)=1.23, *p=*.226.

To check for significant differences in calibration between the pre-test and the post-test by research groups, a two-way repeated-measures analysis of variance (points in time: pre and post) x 2 (research groups: “Meta\_read” and “Control\_read”) was performed. In this analysis, a principal significant difference was found in calibration over time: *F*(1,59)=6.555, *p*<.05, =.100. A significant time x group interaction effect was found as well: *F*(1,59)=4.974, *p*<.05, =.078.

Figure 4 presents the interaction effect graphically.

Insert Figure 4 here

The averages reported in Figure X [4?] and the simple-effect analyses reveal a significant difference between the tests in the “Meta\_read” group, *F*(1,30)=11.316, *p*<.005, =.274, such that calibration was better (i.e., the difference was smaller) on the post-test (*M*=1.290, *SD*=1.346) than on the pre-test (*M*=2.258, *SD*=1.210). In the “Control\_read” group, no significant differences between the tests were found.

To check for a significant difference in calibration between the research groups for each of the six tests, an independent-samples t-test was performed, with the average of the six absolute differences as the dependent variable and the research group as the independent variable. The results of the t-test revealed a significant difference between the research groups in average calibration, *t*(55.612)=2.923, *p*=.005, such that the disparities in the “Meta\_read” group (*M*=1.945, *SD*=0.919) were significantly lower than those in the “Control\_read” group (*M*=2.799, *SD*=1.393).

To examine the effect of metacognitive support on calibration among participants who achieved at different levels, the participants were divided into two groups on the basis of the median pre-test score (a score of 4 out of 7). Those who surpassed the median score (*n*=35) were placed in the first group (high achievers) and those scoring at or below the median (*n*=30) were assigned to the second group (low achievers).

To check for differences in the effect, an independent-samples t-test was performed among members of each of these groups. The results revealed significant calibration differences in the latter group (low achievers) between those who received metacognitive support (“Meta\_read”) and those who did not (“Control\_read”): *t*(33)=2.43, *p*=.021. Participants who received support were more accurate in their judgments (*M=*2.38, *SD=*1.08) than were those not receiving support (*M=*3.43, *SD=*1.38). Among those in the first group (high achievers), the differences were not significant. The differences are shown in Figure 3 [5?].

Figure 5: Calibration between low achievers and high achievers in the research groups.

Insert Figure 5 here

**Resolution**

Resolution denotes the extent to which judgment differentiates between higher levels and lower levels of actual achievements. A common way of measuring resolution is Goodman and Kruskal's gamma (Nelson, 1984; Thiede et al., 2003). In the current study, too, the gamma was calculated for each participant. To make sure their judgments were made rationally as opposed to randomly, a single-sample t-test was performed compared with 0. In this test, the average gamma was found to be significantly different from 0 (*p*<.001), indicating that the participants applied intelligent judgment. In this study, weak correlations were found in both research groups (“Meta\_read”: *M*=0.32, *SD*=0.45, “Control\_read”: *M*=0.269, *SD*=0.47), with no significant differences.

To check for a difference in resolution between low-achieving and high-achieving participants, a t-test was performed. It found a significant difference in gamma between the achievement groups (*M*=-.066, *SD*=0.591 vs. *M*=.301, *SD*=0.545, respectively). Moreover, a significant relation between calibration and resolution was found among low achievers: the stronger the resolution, the smaller the average disparities (r=-.458, p=.024).

**3.3 Discussion**

The results of this study indicate that there was meaningful improvement in reading literacy achievements in learning in a digital environment among students who receive support by being asked metacognitive questions. These findings reinforce previous studies examining metacognitive support, and expand on earlier findings (Author 1, 2013; Valencia-Vallejo et al., 2019). The study expands on its predecessors by shedding light on metacognitive processes that take place in literacy studies in a digital environment. It describes the process of improvement that the research group underwent: the six-session intervention program was demonstrated to improve literacy achievements from one session to the next, using metacognitive support. There was an upward trend, though not significant, in participants’ achievements from test to test.

The study introduced another variable related to literacy achievements: accuracy of judgment.

Participants were asked to judge their learning by predicting their achievements on a comprehension test. As in other studies (Bjork et al., 2013), it was found here that calibration is a metacognitive process associated with effective learning and strong achievements. The significant relationship found between strong achievements and calibration is reflected in the absolute disparity between prediction and achievement, such that the smaller the disparity, the better the achievements. This significant connection was present in both the experimental group and the control group, and among high achievers and low achievers alike.

Calibration is an especially important element in the context of learning in a digital environment, where inaccurate judgments of learning are typical (Ackerman & Goldsmith, 2011). Accordingly, this study sought to introduce a metacognitive support program that would improve calibration. The findings prove that a support program that includes asking metacognitive questions does have a positive effect on calibration. In the six-session intervention program, it was found that metacognitive support induced an improvement in calibration from one meeting to the next. The participants’ judgment became increasingly accurate, such that the gap between their evaluation and their actual achievements narrowed from test to test.

In an additional finding, the effect of metacognitive support on judgment accuracy differed among participants at different levels of achievement. Low-achieving participants who received support showed significantly stronger calibration than did low achievers who did not receive support. It follows that metacognitive support of digital learning is an immensely valuable strategy for low-achieving students in particular. This conclusion reinforces other studies that demonstrate the efficacy of metacognitive support in improving underachieving students’ achievements (Raes et al., 2012; Zohar & David, 2008; Zohar & Peled, 2008).

In contrast to the improvement in calibration, the current study found low resolution values in both research groups. Namely, participants had trouble differentiating between texts that they had learned and understood well and those they had not learned well and required further study. This finding recurs in many studies on judgment of comprehension among young students (Pieger et al., 2016; Redford et al., 2012). The lengthy lapse of time between readings of texts (a week or longer) may have degraded the participants’ resolution by making it hard for them to determine which texts they had learned well and which they had not.

To test the effect of the metacognitive support from a broader perspective, an additional trial took place, Study 2, in which the effect of metacognitive support on learning texts in the context of mathematical literacy was examined. Mathematical literacy has characteristics that differ from those of reading literacy. Mathematical texts, for example, are shorter than reading-literacy texts in terms of word count; they also demand comprehension of mathematical symbols and language. Accordingly, the purpose of Study 2 was to pinpoint the contribution of metacognitive support to accuracy of judgment in learning mathematical literacy in a digital environment.

1. **Study 2**

**4.1 Method**

**4.1.1 Participants**

Seventy-two fifth-graders (age 10–11, 45.8% girls) took part in this study. They attended two different schools that had been invited to take part in the study and gave their consent. The schools were of similar socioeconomic status as defined by the Israel Central Bureau of Statistics (2019). Students who defined themselves as recent immigrants were omitted from the sample because their difficulties in reading comprehension might be based on the challenges of reading in a foreign language.

The students were divided randomly into two groups. Group 1, “Meta\_math” (*n*=33, 47.1% girls) engaged in lessons on mathematical literacy and received metacognitive support in the form of questions integrated into the mathematical text. Group 2, “Control\_math” (*n*=39, 44.7% girls) engaged in lessons on mathematical literacy without receiving support.

**4.1.2 The Intervention and Judgment Program**

The intervention program in Study 2 was identical to that in Study 1, comprised of six texts that examined mathematical literacy, tailored to the fifty-grade curriculum (RAMA, 2016). As in Study 1, four metacognitive self-questions following the IMPROVE model (Mevarech & Kramarski) were integrated into the texts.

**4.1.3 Research Tools**

Accuracy in judging comprehension was measured as it was measured in Study 1; see Section 3.1.3.

*Evaluating Mathematical Literacy Achievements*

The participants’ mathematical literacy was measured by having them read, study, and be tested on six texts that were developed for the purposes of the study. The texts were around 100 words long on average. Each text included a situation from daily life that poses a mathematical problem; the participants were asked to solve the problem and answer a seven-item multiple-choice test by applying mathematical procedures. The texts and the tests were thoroughly vetted for clarity and coherence by a “group of experts in teaching mathematics”, such that they could constitute a tool to measure students’ mathematical literacy.

A reliability check for internal consistency (Cronbach’s Alpha) across the full set of questions on all the tests (42 items) yielded an α=0.804 reliability coefficient. On the basis of this coefficient, a general index score on the comprehension test was calculated for each participant by averaging all 42 items in total. A separate score for each test was also calculated. Each score was arrayed on a 0–7 scale; the higher the score, the higher the level of achievement.

**4.1.4 Procedure**

The preparation and performance stages of this study were identical to those in Study 1; see Section 3.1.4.

**4.2 Results**

The participants’ mathematical literacy achievements, divided by research group, are shown in Table 4.

Insert Table 4 here

The pre-test given to the participants did not include metacognitive support. To rule out the possibility of variance in the research groups’ achievements, an independent-samples t-test was performed. Indeed, no difference in achievements was found between the “Control\_read” group and the “Meta\_read” groups: *t*(68)=.29, *p*=.769).

To check for a significant difference in achievements between the pre-test and the post-test within each research group, a two-way repeated-measures analysis of variance was carried out (points in time: pre- and post-) x 2 (research groups: “Meta\_math” and “Control\_math”). In this analysis, a principal significant difference was found for time: *F*(1,67)=28.228, *p*<.001, =.296. A significant time x group interaction effect was also found: *F*(1,67)=11.102, *p*<.005, =.142.

Figure 7 presents the interaction effect graphically:

Insert Figure 7 here

The averages presented in Figure 7 and the simple-effect analysis reveal a significant difference between the tests in the “Meta\_math” group, meaning that achievements on the post-test surpassed those on the pre-test: *F*(1,32)=40.245, p<.001, =.557. In the “Control\_math” group, no significant achievement differences between the tests were found.

To check for significant differences in literacy achievements between the six tests, an independent-samples t-test was performed, with the average of the six test scores as the dependent variable and the research group as the independent variable. The results of the t-test show a significant difference between the research groups in average scores: *t*(70)=3.413, p<.005, with the achievements of the “Meta\_math” group (*M*=4.136, *SD*=1.053) significantly surpassing those of the “Control\_read” group (*M*=3.284, *SD*=1.063).

**Judgment of Comprehension**

Figure 6 presents the judgments and achievements on each of the six tests, divided by research group.

Insert Figure 6 here

**Absolute Accuracy/Calibration**

One aspect of accurate judgment is calibration, defined as the absolute difference between judgment before the test and actual achievements on the test. Table 5 presents the participants’ calibrations on the six tests, divided by research group.

To examine the relationship between calibration and reading-literacy [mathematical-literacy?] achievements, the correlation between the absolute disparity and achievement was measured. Here a significant negative relationship was found between the absolute gap and the achievements (*r*=-.718, *p*<0.001), i.e., the smaller the disparity was—i.e., the better the calibration—the higher the achievements were.

The participants’ pre-test did not include metacognitive support. To make sure there was no variance in calibration between the research groups, an independent-samples t-test was performed. Indeed, the “Control\_math” and the “Meta\_math” groups showed no difference in calibration: *t*(68)=.639, *p=*.525.

To check for a significant difference in calibration between the pre-test and the post-test within each research group, a two-way repeated-measures analysis of variance (points in time: pre and post) x 2 (research groups: “Meta\_math” and “Control\_math”) was performed. In this analysis, a significant difference was found in calibration over time: *F*(1,67)=4.908, *p*<.05, =.068. Also, a significant time x group interaction effect was found: *F*(1,67)=4.104, *p*<.05, =.058.

Figure 8 presents the interaction effect graphically.

Insert Figure 8 here

The averages reported in Figure X [8?] and the simple-effect analyses reveal a significant difference between the tests in the “Meta\_math” group, *F*(1,32)=8.067, *p*<.01, =.201, such that calibration was better on the post-test (*M*=1.515, *SD*=1.752) than on the pre-test (*M*=2.757, *SD*=1.904). In the “Control\_math” group, no significant differences between the tests were found.

To check for a significant difference in calibration in the research groups for each of the six tests, an independent-samples t-test was performed with the average of the six absolute differences as the dependent variable and the research group (“Meta\_math” or “Control\_math”) as the independent variable. The results of the t-test show a significant difference between the research groups in average calibration, *t*(70)=3.359, p=.001, such that the disparities in the “Meta\_math” group (*M*=2.098, *SD*=0.859) were significantly lower than those of the “Control\_math” group (*M*=2.851, *SD*=1.029).

To examine the effect of metacognitive support on calibration among participants with different achievement levels, the participants were divided into two groups based on the median of their average scores on the pre-test. Participants who scored higher than the median score (*n*=36) were placed in the first group (high achievers) and those scoring at or below the median (*n*=36) were assigned to the second group (low achievers).

To test differences in the effect, an independent-samples t-test was performed among members of each of the two groups. The results show that in the former group (the high achievers) there were significant differences in calibration between those who received metacognitive support (“Meta\_math”) and those who did not (“Control\_math”): *t*(34)=3.42, *p*=.002. Participants who received support were more accurate in their judgments (*M=*1.53, *SD=*0.73) than did those not receiving support (*M=*2.4, *SD=*0.78). Figure 3 presents these differences. Among the low achievers, there were significant differences in calibration between participants who received metacognitive support (“Meta\_math”) and those who did not (“Control\_math”), *t*(34)=2.63, *p*=.013. Those who received support judged their comprehension more accurately (*M=*2.61, *SD=*0.54) than did those who did not receive support (*M=*3.36, *SD=*1.05).

**Resolution**

In Study 2, as in Study 1, a single-sample gamma t-test compared with 0 was carried out. The average gamma was found to be significantly different from 0 (*p*<.007), indicating that the participants applied sensible judgment. In this study, weak correlations were found in both research groups (“Meta\_math”: *M*=.235, *SD*=.601, “Control\_math”: *M*=.167, *SD*=.523), with no significant differences.

**4.3 Discussion**

The results of this study indicate that there was significant improvement in mathematical literacy achievements in learning in a digital environment among students who received support by being asked metacognitive questions. As in Study 1, Study 2 illuminates the process of improvement that the research group underwent. The six-session intervention program showed that metacognitive support induced improvement in literacy achievements from each session to the next. There was an upward trend, though not significant, in participants’ achievements from one test to the next.

As in Study 1, described above, and additional studies (Bjork et al., 2013), it was found in Study 2 that calibration is a metacognitive process associated with effective learning and strong achievements. The significant relationship that was discovered between strong achievements and calibration is reflected in the absolute disparity between prediction and achievement, such that the smaller the disparity, the better the achievements.

The findings of this study demonstrate that the support program was also effective in enhancing calibration. Over the course of the six-session intervention, metacognitive support led to an improvement in calibration from one session to the next. The accuracy of the participants’ judgment improved such that the correspondence of judgment to actual achievements increased from test to test.

In contrast to the improvement in calibration, Study 2 yielded low values of resolution among both research groups. The participants found it hard to differentiate between texts that they had learned and comprehended well and those learned poorly that required further study. This finding appears in many studies on comprehension judgment among young students (Pieger et al., 2016; Redford et al, 2012). The lengthy lapse of time between readings of texts (a week or more) may have degraded resolution by making it hard for the participants to determine which texts they had learned well and which they had not.

1. **General Discussion**

Metacognitive support, through asking participants metacognitive questions, was found to successfully improve calibration in both of the studies conducted in the present research: Study 1, which examined reading literacy, and Study 2, which examined mathematical literacy. As shown in other studies, calibration leads to effective learning and strong achievements. In both studies reported here, a relationship between calibration and high achievements was found; the intervention program providing metacognitive support improved calibration and, in turn, enhanced literacy achievements in reading and mathematics.

In contrast, in both studies conducted here, no improvement in resolution was found among the groups that received metacognitive support. Namely, participants found it difficult to differentiate between texts that they had learned and comprehended well and those that they had not learned well, which required further study. The lengthy lapse of time between readings of texts (a week or more) may have degraded the participants’ resolution by making it hard for them to determine which texts they had learned well and which they had not.

There was a difference between Study 1 and Study 2 with regards to the effect of metacognitive support on calibration among students at different levels of achievement. In Study 1, the intervention program was found to be especially effective among low-achieving participants. Low-achieving participants who received metacognitive support showed a larger significant improvement in calibration than did high achievers. In Study 2, no such difference was found; the improvement of calibration did not vary among students at different levels of achievement.

This can perhaps be explained by the cognitive load associated with reading a “wordy” text—an effect manifested only in Study 1, which examined reading literacy. Low-achieving participants found it more difficult to cope with this load than did higher achievers, but those with low achievements who received metacognitive support managed to cope with the load and improve their calibration. In Study 2, in contrast to Study 1, no cognitive load was created because the texts were short; therefore, there was no difference in the improvement of calibration among participants at different levels of achievement.

**Limitations of the Study and Directions for Future Research:**

This study had several noteworthy limitations. First, Study 1 was based on reading texts of different kinds. The research framework did not allow the researchers to determine whether the effect of the metacognitive support on judgment accuracy and achievements depended on the genre of the text. In view of studies examining adult readers that found an association between text genre and accuracy of judgment (Delgado et al., 2018; Golke, 2019), the role of text genre in the context of judgment accuracy among young readers should be tested in future research. Second, the research matrix comprised six stages, where each stage included reading a lengthy text followed by testing; it took roughly an hour and was carried out once every week or two. This interval was determined in order to test the efficacy of the intervention program over time. Accordingly, much time passed between readings. This may have made it hard to differentiate between texts that were learned well and those that were not, thus creating difficulty in resolution. As in studies that calculated resolution on the basis of reading different texts in one sitting (Dunlosky & Lipko, 2007), future research should be designed such that several short texts would be read in one sitting along with judgment of comprehension. Third, the intervention program included metacognitive questions that were integrated into and presented within the text. The present research framework offered no possibility of animating the questions. If the questions had been presented in an active and animated manner as the reading progressed, they might have been more effective; this should to be investigated in future research. Finally, the text-comprehension tests included multiple-choice questions but not open-ended ones. In reference to studies that noted the drawbacks of tests that lack open-ended questions (Ozuru et al., 2013), courseware in future research should present metacognitive questions in an active and animated manner.

The results of this study hold great importance for educational systems. Reading and learning in digital environments have become common tasks for learners of all ages. The present study proposes an intervention program that can help to improve metacognitive learning processes, comprehension judgment, and achievements in reading literacy.

**Figure 1**

*Examples of a Reading Literacy Text by the "Literacy - That's Me!" Courseware*

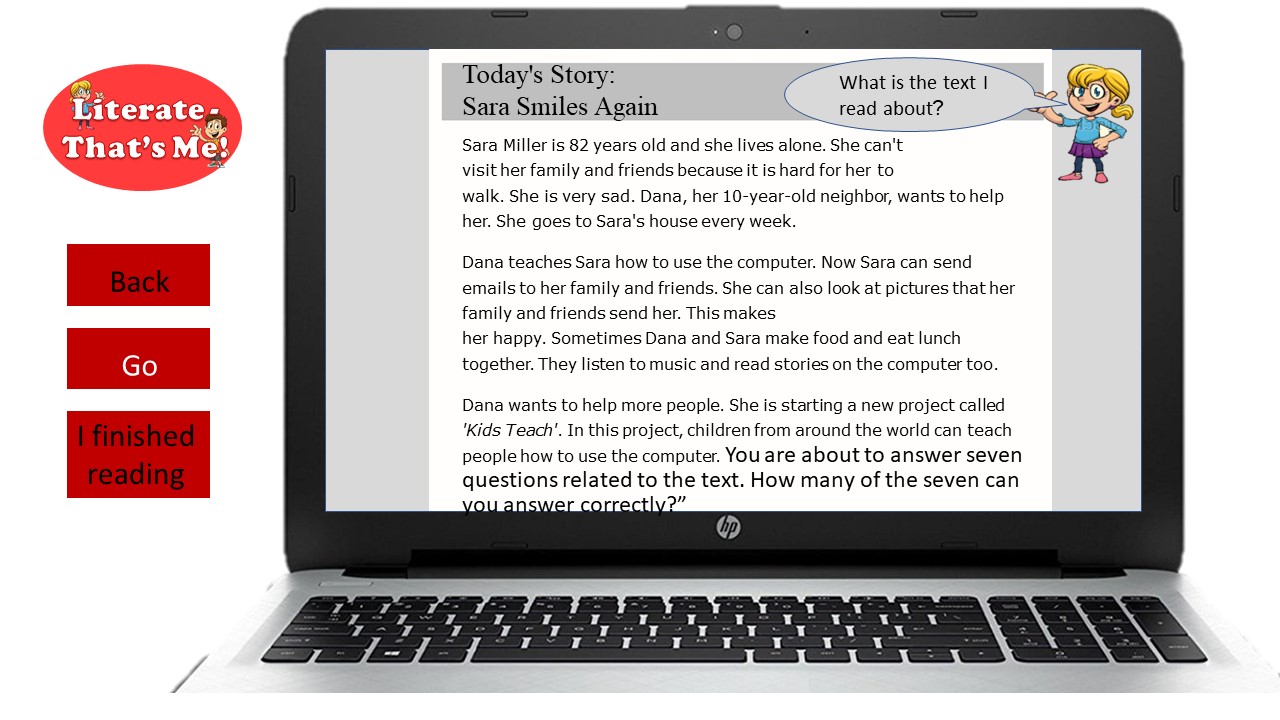
תמונה שמכילה אלקטרוניקה, מחשב

התיאור נוצר באופן אוטומטי

Note. This Text Used in Study 1. The "Literacy - That's Me!" software introduces the text, with the option of continuing to read, reverse or finish.

**Figure 2**

*Examples of Metacognitive Question by the "Literacy - That's Me!" Software*



Note. For the Meta\_read group, the courseware introduces a metacognitive question.

**Figure 3**

*Judgment of Learning—Prediction-of-Performance Question*

תמונה שמכילה מחשב, אלקטרוניקה, צילום מסך

התיאור נוצר באופן אוטומטי

Note. The prediction-of-performance question: You are about to answer seven questions related to the text. How many of the seven can you answer correctly?

**Figure 4**

*Examples of Question in Literacy Test*

תמונה שמכילה אלקטרוניקה, מחשב

התיאור נוצר באופן אוטומטי

Note. The "Literacy - That's Me!" Software introduces a literacy question. The student was asked to choose one correct answer.

**Table 1**

*Summary of Actions Taken at Each Stage of the Study*

|  |  |  |
| --- | --- | --- |
| **Stage of performance** | **Action taken** | **Instructions** |
| **Stage A** | Activating the courseware | Participants are asked to input a personal code. They are allowed to mark in their notebooks anything that may help them understand and remember what they have read. |
| **Stage B** | Reading a text | Metacognitive questions are integrated into texts read by the “Meta\_read” research groups; participants are asked to answer them in writing in their notebooks. |
| **Stage C** | Judgment of comprehension on a scale | Answer the question: “How many of the seven [questions] will you be able to answer correctly?” by sliding a cursor across the screen. |
| **Stage D** | The option of re-reading | Answer the question: “Would you like to read the text again?” If the participant wishes to re-read the text, he or she will be asked to re-judge his or her comprehension. |
| **Stage E** | Being tested on the text | Answer seven questions by choosing one of four possible responses. |
| **Stage F** | The option of correcting responses | Answer the question: “Would you like to change one of the responses that you selected?” |

**Table 2**

*Student Achievements in Reading Literacy by Research Groups*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Control\_read | | Meta\_read | |  |
| *SD* | *M* | *SD* | *M* |  |
| 2.07 | 3.44 | 2.05 | 3.43 | Test1 |
| 1.69 | 2.94 | 1.36 | 3.59 | Test2 |
| 1.68 | 2.82 | 1.68 | 4.10 | Test3 |
| 1.59 | 3.66 | 1.36 | 4.16 | Test4 |
| 1.68 | 3.67 | 2.24 | 4.45 | Test5 |
| 2.44 | 3.10 | 1.86 | 4.84 | Test6 |

**Table 2**

*Calibration Indices in Reading Literacy by Research Groups*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Control\_read | | Meta\_read | |  |
| *SD* | *M* | *SD* | *M* |  |
| 1.98 | 2.53 | 0.97 | 1.46 | Test1 |
| 1.83 | 3.36 | 1.56 | 2.10 | Test2 |
| 1.92 | 3.27 | 1.23 | 1.96 | Test3 |
| 1.82 | 2.81 | 1.51 | 1.93 | Test4 |
| 1.62 | 2.30 | 1.44 | 1.89 | Test5 |
| 2.30 | 2.54 | 1.57 | 1.75 | Test6 |

**Table 4**

*Student Achievements in Math Literacy by Research Groups*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Control\_math | | Meta\_math | |  |
| *SD* | *M* | *SD* | *M* |  |
| 1.59 | 2.97 | 1.75 | 3.09 | Test1 |
| 1.54 | 3.32 | 1.88 | 4.00 | Test2 |
| 1.65 | 3.11 | 1.51 | 3.94 | Test3 |
| 1.55 | 3.42 | 1.51 | 4.03 | Test4 |
| 1.53 | 3.33 | 1.77 | 4.56 | Test5 |
| 1.90 | 3.47 | 1.70 | 5.27 | Test6 |

**Table 3**

*Calibration Indices in Math Literacy by Research Groups*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Control\_math | | Meta\_math | |  |
| *SD* | *M* | *SD* | *M* |  |
| 1.62 | 3.02 | 1.90 | 2.75 | Test1 |
| 1.70 | 2.78 | 1.52 | 1.84 | Test2 |
| 1.64 | 3.16 | 1.44 | 2.27 | Test3 |
| 1.58 | 2.36 | 1.78 | 2.28 | Test4 |
| 1.70 | 2.86 | 1.31 | 1.82 | Test5 |
| 2.07 | 2.97 | 1.75 | 1.51 | Test6 |

**Figure 5**

*Pre-Test and Post-Test Achievements*

**Figure 6**

Judgments and Achievements on the Tests

**Figure 7**

*Judgments and Achievements* *on* *the Tests*

**Figure 8**

*Pre-Test and Post-Test Calibration*

**Figure 9**

*Calibration among High Achievers and Low Achievers*

**Figure 10**

*Pre-Test and Post-Test Achievements*

**Figure 10**

*Pre-Test and Post-Test Calibration*

# Bibliography

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied*, *17*(1), 18–32. https://doi.org/10.1037/a0022086

Ackerman, R., & Lauterman, T. (2012). Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure. *Computers in Human Behavior*, *28*(5), 1816–1828. https://doi.org/10.1016/j.chb.2012.04.023

Annisette, L. E., & Lafreniere, K. D. (2017). Social media, texting, and personality: A test of the shallowing hypothesis. *Personality and Individual Differences*, *115*, 154–158. https://doi.org/10.1016/J.PAID.2016.02.043

Baars, M., van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of Problem Solving after Worked Example Study on Primary School Children’s Monitoring Accuracy. *Applied Cognitive Psychology*, *28*(3), 382–391. https://doi.org/10.1002/acp.3008

Benedetto, S., Drai-Zerbib, V., Pedrotti, M., Tissier, G., & Baccino, T. (2013). E-Readers and Visual Fatigue. *PLoS ONE*, *8*(12), e83676. https://doi.org/10.1371/journal.pone.0083676

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). *Self-Regulated Learning: Beliefs, Techniques, and Illusions*. https://doi.org/10.1146/annurev-psych-113011-143823

Bol, L., Hacker, D. J., O’Shea, P., & Allen, D. (2005). The Influence of Overt Practice, Achievement Level, and Explanatory Style on Calibration Accuracy and Performance. *The Journal of Experimental Education*, *73*(4), 269–290. https://doi.org/10.3200/JEXE.73.4.269-290

Boström, L., & Lassen, L. M. (2006). Unraveling learning, learning styles, learning strategies and meta‐cognition. *Education + Training*, *48*(2/3), 178–189. https://doi.org/10.1108/00400910610651809

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Carretti, B., Caldarola, N., Tencati, C., & Cornoldi, C. (2014). Improving reading comprehension in reading and listening settings: The effect of two training programmes focusing on metacognition and working memory. *British Journal of Educational Psychology*, *84*(2), 194–210. https://doi.org/10.1111/bjep.12022

Central Bureau of Statistics. (2019).

Chou, I.-C. (2016). Reading for the purpose of responding to literature: EFL students’ perceptions of e-books. *Computer Assisted Language Learning*, *29*(1), 1–20. https://doi.org/10.1080/09588221.2014.881388

Csapó, B., Ainley, J., Bennett, R. E., Latour, T., & Law, N. (2012). Technological Issues for Computer-Based Assessment. In *Assessment and Teaching of 21st Century Skills* (pp. 143–230). Springer Netherlands. https://doi.org/10.1007/978-94-007-2324-5\_4

Dahan Golan, D., Barzillai, M., & Katzir, T. (2018). The effect of presentation mode on children’s reading preferences, performance, and self-evaluations. *Computers & Education*, *126*, 346–358. https://doi.org/10.1016/J.COMPEDU.2018.08.001

Daniel, D. B., & Woody, W. D. (2013). E-textbooks at what cost? Performance and use of electronic v. print texts. *Computers & Education*, *62*, 18–23. https://doi.org/10.1016/J.COMPEDU.2012.10.016

de Boer, H., Donker, A. S., Kostons, D. D. N. M., & van der Werf, G. P. C. (2018). Long-term effects of metacognitive strategy instruction on student academic performance: A meta-analysis. *Educational Research Review*, *24*, 98–115. https://doi.org/10.1016/J.EDUREV.2018.03.002

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*, *25*, 23–38. https://doi.org/10.1016/J.EDUREV.2018.09.003

Dennis, A. R., Abaci, S., Morrone, A. S., Plaskoff, J., & McNamara, K. O. (2016). Effects of e-textbook instructor annotations on learner performance. *Journal of Computing in Higher Education*, *28*(2), 221–235. https://doi.org/10.1007/s12528-016-9109-x

Duncan, L. G., McGeown, S. P., Griffiths, Y. M., Stothard, S. E., & Dobai, A. (2015). Adolescent reading skill and engagement with digital and traditional literacies as predictors of reading comprehension. *British Journal of Psychology*, *107*(2), 209–238. https://doi.org/10.1111/bjop.12134

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension. *Current Directions in Psychological Science*, *16*(4), 228–232. https://doi.org/10.1111/j.1467-8721.2007.00509.x

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students’ learning and retention. *Learning and Instruction*, *22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, *98*(1), 37–56. https://doi.org/10.1016/S0001-6918(97)00051-6

Ellis, A. K., Denton, D. W., & Bond, J. B. (2014). An Analysis of Research on Metacognitive Teaching Strategies. *Procedia - Social and Behavioral Sciences*, *116*, 4015–4024. https://doi.org/10.1016/j.sbspro.2014.01.883

Forbes, K., & Fisher, L. (2015). The impact of expanding advanced level secondary school students’ awareness and use of metacognitive learning strategies on confidence and proficiency in foreign language speaking skills. *The Language Learning Journal*, 1–15. https://doi.org/10.1080/09571736.2015.1010448

Golke, S., Hagen, R., & Wittwer, J. (2019). Lost in narrative? The effect of informative narratives on text comprehension and metacomprehension accuracy. *Learning and Instruction*, *60*, 1–19. https://doi.org/10.1016/j.learninstruc.2018.11.003

Jeong, H. (2012). A comparison of the influence of electronic books and paper books on reading comprehension, eye fatigue, and perception. *The Electronic Library*, *30*(3), 390–408. https://doi.org/10.1108/02640471211241663

Kim, Y. H. (2016). Early childhood educators’ meta-cognitive knowledge of problem-solving strategies and quality of childcare curriculum implementation. *Educational Psychology*, *36*(4), 658–674. https://doi.org/10.1080/01443410.2014.934661

Kong, Y., Seo, Y. S., & Zhai, L. (2018). Comparison of reading performance on screen and on paper: A meta-analysis. *Computers & Education*, *123*, 138–149. https://doi.org/10.1016/J.COMPEDU.2018.05.005

Kramarski, B., & Kohen, Z. (2017). Promoting preservice teachers’ dual self-regulation roles as learners and as teachers: effects of generic vs. specific prompts. *Metacognition and Learning*, *12*(2), 157–191. https://doi.org/10.1007/s11409-016-9164-8

Lauterman, T., & Ackerman, R. (2014). Overcoming screen inferiority in learning and calibration. *Computers in Human Behavior*, *35*, 455–463. https://doi.org/10.1016/J.CHB.2014.02.046

Liu, Z. (2005). Reading behavior in the digital environment Changes in reading behavior over the past ten years. *Journal of Documentation*, *61*(6), 700–712. https://doi.org/10.1108/00220410510632040

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. https://doi.org/10.1016/j.ijer.2012.12.002

McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking Reading Comprehension Instruction: A Comparison of Instruction for Strategies and Content Approaches. *Reading Research Quarterly*, *44*(3), 218–253. https://doi.org/10.1598/RRQ.44.3.1

McMaster, K. L., den Broek, P. van, A. Espin, C., White, M. J., Rapp, D. N., Kendeou, P., Bohn-Gettler, C. M., & Carlson, S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, *22*(1), 100–111. https://doi.org/10.1016/j.lindif.2011.11.017

Mevarech, Z., & Fan, L. (2018). *Cognition, Metacognition, and Mathematics Literacy* (pp. 261–278). https://doi.org/10.1007/978-3-319-66659-4\_12

Mevarech, Z., & Kramarski, B. (1997). Improve: A Multidimensional Method For Teaching Mathematics in Heterogeneous Classrooms. *American Educational Research Journal*, *34*(2), 365–394. https://doi.org/10.3102/00028312034002365

Mevarech, Z., & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students’ mathematical reasoning. *British Journal of Educational Psychology*, *73*(4), 449–471. https://doi.org/10.1348/000709903322591181

Michalsky, T. (2013). Integrating Skills and Wills Instruction in Self-Regulated Science Text Reading for Secondary Students. *International Journal of Science Education*, *35*(August), 1846–1873. https://doi.org/10.1080/09500693.2013.805890

Michalsky, T., Mevarech, Z., & Haibi, L. (2009). Elementary School Children Reading Scientific Texts: Effects of Metacognitive Instruction. *The Journal of Educational Research*, *102*(5), 363–376. https://doi.org/10.3200/JOER.102.5.363-376

Mihalca, L., Mengelkamp, C., & Schnotz, W. (2017). Accuracy of metacognitive judgments as a moderator of learner control effectiveness in problem-solving tasks. *Metacognition and Learning*, 1–23. https://doi.org/10.1007/s11409-017-9173-2

Morineau, T., Blanche, C., Tobin, L., & Guéguen, N. (2005). The emergence of the contextual role of the e-book in cognitive processes through an ecological and functional analysis. *International Journal of Human-Computer Studies*, *62*(3), 329–348. https://doi.org/10.1016/j.ijhcs.2004.10.002

Mueller, P. A., & Oppenheimer, D. M. (2014). The Pen Is Mightier Than the Keyboard. *Psychological Science*, *25*(6), 1159–1168. https://doi.org/10.1177/0956797614524581

Muijselaar, M. M. L., Swart, N. M., Steenbeek-Planting, E. G., Droop, M., Verhoeven, L., & de Jong, P. F. (2017). Developmental Relations Between Reading Comprehension and Reading Strategies. *Scientific Studies of Reading*, *21*(3), 194–209. https://doi.org/10.1080/10888438.2017.1278763

Murphy, P. ., Long, J. ., Holleran, T. ., & Esterly, E. (2003). Persuasion online or on paper: a new take on an old issue. *Learning and Instruction*, *13*(5), 511–532. https://doi.org/10.1016/S0959-4752(02)00041-5

Nelson, T., & Narens, L. (1990). METAMEMORY: A THEORETICAL FRAMEWORK AND NEW FINDINGS. *Psychology of Learning and Motivation*, *26*, 125–173. http://www.imbs.uci.edu/~lnarens/1990/Nelson&Narens\_Book\_Chapter\_1990.pdf

OECD. (2018). *PISA 2021 MATHEMATICS FRAMEWORK (FIRST DRAFT)*. https://www.upc.smm.lt/naujienos/smm/penkiolikmeciu-matematinis-rastingumas/GB-2018-4-PISA-2021-Mathematics-Framework-First-Draft.pdf

Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology*, *67*(3), 215–227. https://doi.org/10.1037/a0032918

Pellas, N. (2014). The influence of computer self-efficacy, metacognitive self-regulation and self-esteem on student engagement in online learning programs: Evidence from the virtual world of Second Life. *Computers in Human Behavior*, *35*, 157–170. https://doi.org/10.1016/j.chb.2014.02.048

Pfost, M., Dörfler, T., & Artelt, C. (2013). Students’ extracurricular reading behavior and the development of vocabulary and reading comprehension. *Learning and Individual Differences*, *26*, 89–102. https://doi.org/10.1016/J.LINDIF.2013.04.008

Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency – Does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction*, *44*, 31–40. https://doi.org/10.1016/j.learninstruc.2016.01.012

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

Pishghadam, R., & Khajavy, G. H. (2013). Intelligence and metacognition as predictors of foreign language achievement: A structural equation modeling approach. *Learning and Individual Differences*, *24*, 176–181. https://doi.org/10.1016/j.lindif.2012.12.004

Raes, A., Schellens, T., De Wever, B., & Vanderhoven, E. (2012). Scaffolding information problem solving in web-based collaborative inquiry learning. *Computers and Education*, *59*(1), 82–94. https://doi.org/10.1016/j.compedu.2011.11.010

RAMA - The National Authority for Measurement & Evaluation in Education. (2009). *Medida Besherut Ha’lemida [Assessment for Learning]*.

RAMA - The National Authority for Measurement & Evaluation in Education. (2016). *Meitzav 2006 [School Growth and Efficiency Measures 2006]*.

Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction*, *22*(4), 262–270. https://doi.org/10.1016/J.LEARNINSTRUC.2011.10.007

Sidi, Y., Ophir, Y., & Ackerman, R. (2016). Generalizing screen inferiority - does the medium, screen versus paper, affect performance even with brief tasks? *Metacognition and Learning*, *11*(1), 15–33. https://doi.org/10.1007/s11409-015-9150-6

Sidi, Y., Shpigelman, M., Zalmanov, H., & Ackerman, R. (2017). Understanding Metacognitive Inferiority on Screen by Exposing Cues for Depth of Processing. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2017.01.002

Singer, L. M., & Alexander, P. A. (2017). Reading Across Mediums: Effects of Reading Digital and Print Texts on Comprehension and Calibration. *The Journal of Experimental Education*, *85*(1), 155–172. https://doi.org/10.1080/00220973.2016.1143794

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of Children’s Knowledge and Regulation of Cognition. *Contemporary Educational Psychology*, *27*(1), 51–79. https://doi.org/10.1006/ceps.2001.1091

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66

UNESCO. (2004). *The Plurality of literacy and its Implications for Policies and Programmes: UNESCO Position Paper*. https://books.google.co.il/books/about/The\_Plurality\_of\_Literacy\_and\_Its\_Implic.html?id=awhPswEACAAJ&redir\_esc=y

Valencia-Vallejo, N., López-Vargas, O., & Sanabria-Rodríguez, L. (2019). Effect of a metacognitive scaffolding on self-efficacy, metacognition, and achievement in e-learning environments. *Knowledge Management and E-Learning*, *11*(1), 1–19. https://doi.org/10.34105/j.kmel.2019.11.001

Winer, D. (2018). *Israel Country Report on ICT in Education*. http://meyda.education.gov.il/files/Owl/Hebrew/education-plans.pdf

Winne, P. H., & Hadwin, A. F. (2013). *nStudy: Tracing and Supporting Self-Regulated Learning in the Internet* (pp. 293–308). Springer, New York, NY. https://doi.org/10.1007/978-1-4419-5546-3\_20

Zohar, A., & David, A. Ben. (2008). Explicit teaching of meta-strategic knowledge in authentic classroom situations. *Metacognition and Learning*, *3*(1), 59–82. https://doi.org/10.1007/s11409-007-9019-4

Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low- and high-achieving students. *Learning and Instruction*, *18*(4), 337–353. https://doi.org/10.1016/j.learninstruc.2007.07.001