Scientific Research Publishing

Psychology, 2022, 13, 1727-1747
https://www.scirp.org/journal/psych
ISSN Online: 2152-7199
ISSN Print: 2152-7180

Confidence and Calibration of Comprehension in Adolescence: Are They Domain-General or Domain-Specific?

Einat Kleider-Tesler, Mirit Barzillai, Tami Katzir

Department of Learning Disabilities, The E.J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Haifa, Israel
Email: einat.klte@gmail.com

How to cite this paper: Kleider-Tesler, E., Barzillai, M., & Katzir, T. (2022). Confidence and Calibration of Comprehension in Adolescence: Are They Domain-General or Domain-Specific? *Psychology, 13,* 1727-1747.
https://doi.org/10.4236/psych.2022.1312108

Received: September 14, 2022
Accepted: November 12, 2022
Published: November 15, 2022

Abstract

Self-regulation of learning is important for reading comprehension. Self-regulation is typically measured by confidence rating before or after a task and calibration calculations. Less is known about the stability of confidence and calibration within and across different tasks. In this study, we examined whether confidence ratings and calibration of adolescence differ by task, skill and time of rating (before, during, and after the task). Eighty-six adolescents took a reading comprehension task and a non-verbal spatial reasoning task (TONI) and were asked to self-evaluate their confidence before, during, and after the tasks. Results within tasks showed that ratings were significantly different as a function of time. Differences were also found across tasks. Consistent with findings in other age groups, good comprehenders were better than poor comprehenders in calibration. Our findings suggest that metacognitive evaluations have domain-specific as well as domain-general aspects and are also influenced by reading comprehension skills.

Keywords

Self-Regulation, Reading Comprehension, Reading Comprehension Difficulties, Adolescents

1. Introduction

Success in reading comprehension tasks, particularly in high school, requires students to employ a range of metacognitive skills (Vadhan & Stander, 1994), such as accurate self-evaluations of their performance and appropriate use of reading strategies (Guthrie, Wigfield, Metsala, & Cox, 1999; Meltzer, Katzir, Miller, Reddy,

& Roditi, 2004). Much of the work on self-evaluations in reading comprehension (RC) has focused on children and adults (Ackerman & Lauterman, 2012; Ehrlich, Remond, & Tardieu, 1999; Kasperski & Katzir, 2013), and less is known about self-evaluation of performance among high school students and whether these evaluations are domain-specific or a general meta-cognitive ability. Studies with adults suggest that individuals who consistently reflect and regulate comprehension processes during reading demonstrate more efficient studying and better comprehension (Thiede, Anderson, & Therriault, 2003; Wiley, Griffin, & Thiede, 2005). Such self-regulation is closely related to the ability to "calibrate", or to accurately evaluate one's own learning (Glenberg & Epstein, 1985; Kasperski & Katzir, 2013; Lin, Moore, & Zabrucky, 2001). The students' ability to evaluate their level of text understanding, called also "Calibration of Comprehension", is tied to reading ability (Lin & Zabrucky, 1998). Poor comprehenders frequently overestimate their abilities and are also less likely to know that their functioning is inaccurate or to understand how close they are to the desired learning (Dinsmore & Parkinson, 2013).

Previous research has found that a variety of factors influence readers' self-evaluation judgments. These include reading skills, metacognitive knowledge, task requirements, and text difficulty (Lin & Zabrucky, 1998). The applicability of these results to adolescents, however, is problematic as most studies investigated self-evaluation and calibration abilities among elementary school children and young adults (Ackerman & Leiser, 2014; Glover, 1989; Klassen, 2007; Kasperski & Katzir, 2013). In addition, most studies examine self-evaluation of the task before, during, or after the task, rather than as multiple evaluations in the process of the task. Therefore, in order to understand the nature of these evaluations in depth, our study focuses on three evaluations measured by confidence ratings—before, during, and after a task. Furthermore, most studies focus on self-evaluation in a specific domain (i.e., reading, math). It remains unclear whether the ability to self-evaluate is related to a specific task or domain. On one hand, some researchers concluded that self-evaluation becomes more stable over time and generalizes across tasks (Kleitman & Stankov, 2001; Nietfeld & Schraw, 2002; Stankov & Crawford, 1996, 1997), whereas others indicated that it is specific to certain test items or tasks rather than being a general evaluation of ability or learning (Hadwin & Webster, 2013). In order to address these discrepancies and inconsistencies in the literature, the main goal of this study was to investigate whether adolescents' self-evaluation and calibration skills are domain-general or specific to a task, while examining them within a task (before, during, and after a task) and *across* tasks (in a reading comprehension vs. spatial reasoning task). We also examined whether these confidence ratings and calibration abilities are tied to reading comprehension skill (good vs. poor comprehenders) among adolescents.

1.1. Self-Regulation and Reading Comprehension

Self-regulated learning has been defined as a student's cognitive and meta-cognitive strategies used to control and regulate learning (Garcia & Pintrich, 1994;

Pintrich, 1988). A process related to self-regulation is called monitoring of comprehension, and it occurs when students check their understanding against some self-set goal. One way of measuring self-regulation in reading comprehension is to ask individuals to rate their confidence as to whether they gave correct responses, using a numerical percentage scale. This is followed by calculating the difference between the confidence rating and actual performance, also called calibration of comprehension (CoC) (de Carvalho Filho & Yuzawa, 2001).

Calibration is most often seen as an integral part of self-regulation, as it reflects the ability to monitor and evaluate performance procedures over time (Kleitman & Stankov, 2001; Pallier, Wilkinson, Danthiir, Kleitman, Knezevic, & Stankov, 2002). 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 (Thiede, Anderson, & Therriault, 2003).

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). Indeed, as children enter adolescence, they become more aware of their own competence, more realistic about task demands (Harter, 1998; Koriat, Ackerman, Adiv, Lockl, & Schneider, 2014) and gain the cognitive capacity to be more self-reflective and aware of their own abilities and difficulties (Stipek, 1998). Hence, adolescents' decline in self-evaluation 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 (Koriat, Ackerman, Lockl, & Schneider, 2009), confidence in performance continues to exceed reality in adolescents (Rogers, Vitacco, Jackson, Martin, Collins, & Sewell, 2002). Although older students' judgments are less exaggerated than those of younger students, they continue to overestimate their own abilities (Falchikov & Boud, 1989) in most domains (Kruger & Dunning, 1999) and throughout adulthood.

Most of the studies that observed confidence ratings did so during the task. In order to deepen our understanding of the process of self-regulation, our study aimed to observe this judgment at three different points in time and see whether it changes in the process of a task.

1.2. Self-Evaluation within Tasks (Confidence Rating before, during, and after a Task)

There are many methods in the literature used to examine students' self-evaluations, such as rating the percentage of items students are expected to answer (Hacker, Bol, Horgan, & Rakow, 2000). In this paper, we use confidence ratings to address how confident a student feels about his or her performance in a given moment at three points in time—before (prediction), during, and after a task. In addition, we use the term calibration to designate the difference between confidence rating and actual performance (de Carvalho Filho & Yuzawa, 2001).

Confidence ratings before the task (predictions of performance) require students to evaluate what they know about a test, how thoroughly they understand the material, and whether they will be able to use their knowledge to optimize performance. Several studies have shown that students are not very accurate at predicting their future test performance after reading texts and before answering the questions (Glenberg, Sanocki, Epstein, & Morris, 1987; Maki & Berry, 1984). Confidence ratings during a task occur when learners is asked to indicate whether they successfully answered a question or achieved a self-set goal (Schraw, 2009). The final self-evaluation measure we examined is confidence ratings after a task. Once students have finished a test, they have complete knowledge of the accuracy of their performance. Their speculations about item difficulty, how much learning occurred, or how much learning was forgotten become more definite judgments about what they knew or did not know in retrospect (Pressley & Ghatala, 1989; Pressley, Levin, Ghatala, & Ahmad, 1987).

We found no studies that compared confidence rating before, during, and after a task in general, and especially not in a single RC task. Most studies compared only self-evaluation before and after a task, and most of this was done with regard to performance rather than as a measure in itself. For example, some researchers indicated that students' confidence ratings after a task, in general, are more accurate than their predictions because they often gain additional self-feedback from taking the test (Devolder, Brigham, & Pressley, 1990; Maki, Jonas, & Kallod, 1994). With regard to performance, calibration of performance and confidence ratings after the task are more accurate than calibration of performance and confidence ratings before the task. This finding indicates that students utilize additional information as a result of performing a test as feedback that enables more precise self-evaluations after a task.

Still, it is not clear whether self-evaluation remains stable within the same tasks. Lin, Moore and Zabrucky (2001) showed that, in undergraduate students, calibration measures (prediction and postdiction) significantly correlated with one another, suggesting that students' ability to calibrate comprehension and performance was stable across measures in the same task. They also found a significant relationship between the ability to calibrate comprehension before the test and the ability to calibrate performance after the test. Thus, metacognitive ability assessed using the calibration paradigms seemed to remain stable across self-evaluation measures in the same task. On the other hand, some studies have shown that participants have poorer confidence judgments before the test compared to their confidence judgments after the test, which are more accurate (Maki, 1995; Maki & Serra, 1992).

To summarize, most studies reviewed students' self-evaluations and calibration at one or two times, mostly compared to performance. Moreover, most studies focused on self-evaluation in one specific domain and only among adult or young populations. Consequently, it is unclear whether these self-evaluations remain stable within tasks and across tasks and whether or not these self-evaluations

are domain-general or specific to a certain task, especially in adolescents.

1.3. Self-Evaluation across Tasks: Are Academic Self-Evaluations and Calibration Domain General or Domain Specific?

Prediction accuracy of academic tests is influenced by how much learning has occurred before the text, the difficulty of the test items, students' familiarity with the content, and prior performance on tests (Benjamin & Bjork, 1996; Glenberg et al., 1987; Maki, 1998; Maki & Serra, 1992). Studies have shown that confidence ratings before the task are related to specific performance and task (Maki, Foley, Kajer, Thompson, & Willert, 1990; Maki & Serra, 1992; Weaver, 1990) and not to general ability. According to self-evaluations after the task, results are inconsistent. Several studies suggest that students' self-evaluations after the task remain relatively stable across tests (Kleitman & Stankov, 2001; Nietfeld & Schraw, 2002; Pallier et al., 2002; Stankov & Crawford, 1996, 1997). However, some studies also indicate that confidence judgments after the task are influenced by characteristics of the task, such as the type of questions. For example, Maki (1995) found that posttest confidence judgments were most consistently accurate for questions tapping details.

While the literature on judgments of confidence before and after the task is very limited, the concept of judgments of confidence during a task has been investigated more broadly across different domains. The question of whether such judgments are stable across different domains has been explored in several ways. There are numerous findings showing strong correlations between confidence ratings from different cognitive tests, indicating that such ratings are domain-general (Kleitman & Stankov, 2001; Nietfeld & Schraw, 2002; Pallier et al., 2002; Stankov & Crawford, 1996, 1997). Moreover, some findings support the notion that confidence judgments become more stable over time and task (Pieschl, 2009). On the other hand, several researchers contend that judgments of confidence are domain-specific (Alexander, Dinsmore, Parkinson, & Winters, 2011). For example, experiences that make learning feel more effortful tend to reduce students' confidence in their learning.

In the reading comprehension domain, we know that accurate evaluation of comprehension involves the use of multiple criteria of the task (e.g., text interestingness, type of test) and personal abilities (effort, reading difficulty, self-image) simultaneously (Lin, Moore, & Zabrucky, 2001; Lin, Zabrucky, & Moore, 1997; Schiefele, 1991). However, work on CoC has focused mainly on young students and adults. In addition, few studies have examined patterns of confidence judgments across two complex tasks from different domains, such as judgments of reading comprehension and judgments of performance on a complex conceptual visual task. Less is known about whether the calibration ability is domain-general and related to reading comprehension skills among adolescents. Do poor comprehenders exhibit overconfidence in reading comprehension tasks only, or are they generally overconfident and thus demonstrate miscalibration across domains?

1.4. Self-Evaluation Calibration among Good and Poor Comprehenders

Calibration has been studied extensively in the literature, however few studies have compared calibration in reading comprehension and calibration in a different field, and specifically with comparison between good and poor comprehenders among adolescents. Studies indicate that confidence judgment after a task is associated with skill. Researchers have found that one factor thought to affect CoC is reading skill. Studies have found that poor comprehenders frequently overestimate their abilities. This tendency has been observed in children (Ehrlich et al., 1999) as well as in adults (Glover, 1989; Klassen, 2007; Maki, Shields, Wheeler, & Zacchilli, 2005). For example, among undergraduate students, higher achievers show less overconfidence than lower achievers in reading comprehension tasks (Ackerman & Leiser, 2014). Kasperski and Katzir (2013) further found that elementary school students who are good comprehenders have high confidence in their achievements, as well as high comprehension calibration scores. In addition, they found significant differences among three groups of readers (high, average, and low), suggesting a relationship between reading skill and confidence ratings. They further found that confidence ratings were normally distributed within each group. Thus, while differences do indeed exist among different profiles of readers, even among high comprehenders there are children who are insecure about their reading abilities, whereas among low comprehenders there are children that are extremely secure about their reading abilities.

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 online monitoring of performance. In contrast, students with higher self-monitoring scores show more accuracy in calibrating comprehension performance (Lin, Moore, & Zabrucky, 2001). Further, as Dinsmore and Parkinson (2013) argue, 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.

Among students with LD (learning disabilities), studies have revealed that LD students were generally better at predictions and calibrating mathematical performance than in writing or reading performance. Still, in the domains of reading and writing, calibration accuracy was low, with students with documented writing difficulties making the least accurate performance predictions (Stolp & Zabrucky, 2009). Thus, on the one hand poor comprehenders show specific areas of academic weakness. On the other, some theorists contend that poor comprehenders have more generalized metacognitive weaknesses (Job & Klas-

sen, 2012). For example, Job and Klassen (2012) found differences between LD and normally achieving (NA) adolescents in calibration. Specifically, while students with LD overestimated their abilities in both spelling and ball-throwing tasks, NA students were more accurate in their evaluations. These findings support the claim that students with LD have a broad metacognitive deficit that affects their accuracy in predicting performance (Kruger & Dunning, 1999; Ruban, McCoach, McGuire, & Reis, 2003).

In conclusion, studies indicate that confidence judgment is associated with reading skill, such that poor comprehenders frequently overestimate their abilities. However, it is still not clear whether such miscalibration is specific to academic tasks or to general ability.

1.5. Conclusions

To summarize, self-evaluation measured as judgment of confidence before, during, and after a task, and calibration, are necessary for academic success (Flavell, 1976). Evidence suggests that self-evaluation and calibration skills may be domain-general (Kleitman & Stankov, 2001, Nietfeld & Schraw, 2002), but they are also influenced by task (Maki, 1995; Hadwin & Webster, 2013) and skill (Ackerman & Leiser, 2014; Ehrlich et al., 1999; Glover, 1989; Klassen, 2007; Maki et al., 2005). Less skilled comprehenders have been shown to be less calibrated than skilled comprehenders, with the majority of studies focusing on adult populations and children (Ehrlich et al., 1999; Kasperski & Katzir, 2013). Thus, little is known about whether the ability to self-evaluate and calibrate is domain-general and stable within and across tasks and whether it is tied to reading comprehension skills, among adolescents.

1.6. Specific Research Questions and Predictions

Our study was guided by the following questions:

1) Is self-evaluation stable throughout task performance?

Do adolescents exhibit a difference in self-evaluation *within* a task (before, during, and after a task of reading comprehension and spatial reasoning)?

2) Are academic self-evaluations and calibration domain-general or task-specific?

Do adolescents exhibit different patterns of self-evaluation and calibration *across* tasks (reading comprehension and a spatial reasoning task)?

3) Are self-evaluations and calibration abilities associated with general reading comprehension skill? Do good and poor comprehenders differ in their self-evaluations and calibration scores?

Based on findings from previous research, we predicted that students' self-evaluations before, during, and after a task would differ within and across verbal and spatial reasoning tasks (Finney & Schraw, 2003). Differences in evaluations within the same task and between tasks will support the argument that confidence judgments are influenced by the nature of the task and its content rather than by general ability across domains. Second, we hypothesized that poor com-

prehenders would overestimate their performance more than their peers (Finney & Schraw, 2003; Klassen, 2007). Comparing calibration and self-evaluations across reading comprehension and perceptual tasks will shed light on the specificity of metacognitive processes.

2. Material and Methods

2.1. Participants

For the purpose of this study, a sample of 86 tenth graders (42 female, 44 male, mean age 15.52 years) were recruited from two high schools in central Israel. Of all participants, 97.7% were born in Israel. All participants were educated in mainstream settings. Seven classes were randomly selected and all students whose parents signed agreement consent form were included in the study.

2.2. Experimental Measures

2.2.1. Cognitive Measures

Personal information questionnaire. Non-identifying personal details were collected from all participants, with regard to their age, gender, mother tongue, place of birth, and date of immigration.

GORT-Reading comprehension. Reading comprehension was measured by a Hebrew version of the Gray Oral Reading Test-Diagnostic (GORT-D; Bryant & Wiederholt, 1991; Prior, Zeltsman-Kulick, & Kazir, 2020). In the present study, seven texts were selected. The first served as an example (suitable for a high school level). Each text was followed by five multiple-choice questions. The questions targeted both factual understanding of the text as well as inferential understanding. Participants read the passages silently and answered the questions in a classroom setting. Reading comprehension scores were calculated based on the number of correct responses. The Hebrew version of the task showed good internal consistency in other studies (Cronbach's alpha = .958).

TONI. The Test of Nonverbal Intelligence-Fourth Edition (TONI-4) served as a non-verbal task. The test aims to measure two components of intelligence: abstract reasoning and problem solving, based on Brown, Sherbenou and Johnsen (2010). The participants were presented with 45 sequences of shapes (Brown et al., 2010), ascending in difficulty level. For each item, participants were asked to choose the next object in the sequence from 4 - 6 possible answers. The number of correct answers was summed (Brown, Sherbenou, & Johnsen, 2010). Reliability was high (Cronbach's alpha = .842).

2.2.2. Metacognition Measures

Aggregated confidence ratings before the task. Before answering each test (reading comprehension and TONI), the participants were asked to predict their performance on the tasks ("How confident are you that you will respond correctly throughout the test"?). We used a 100 point scale ranging from 0 (very unconfident) to 100 (very confident).

Confidence ratings during the task. Itemized confidence ratings for perfor-

mance on 30 reading comprehension questions and for the 45 items of the TONI test were measured for each participant. Reading comprehension confidence scores were measured in accordance with previous research (Kleitman & Stankov, 2001; Schraw & Roedel, 1994; West & Stanovich, 1997). Following each question or item, 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).

Aggregated confidence ratings after the task. Immediately after completing the last test or item participants were asked: "Please indicate how confident you are in your overall performance on this test" (Nietfeld, Cao, & Osborne, 2005). We used a 100 point scale ranging from 0 (very unconfident) to 100 (very confident).

Calibration of Comprehension. This measure consisted of the absolute value of the difference between the confidence rating during the task and performance for each test. First, we calculated the percentage of correct responses (for all items). Then we calculated the difference between mean confidence ratings and performance (Nietfeld, Cao, & Osborne, 2005).

Bias. Bias consisted of the difference between the average confidence rating during the task and average performance scores on each test. Positive scores indicate underconfidence, and negative scores indicate overconfidence. The further the score is from 0, the more biased it is (Nietfeld, Cao, & Osborne, 2005).

2.3. Design and Procedure

Participants were asked to complete a personal information questionnaire. The reading comprehension tests and TONI were administered in a quiet setting in small groups of 10 - 15. Participants received oral and written instructions for each task and were asked to complete sample items. Students completed the tasks independently and were encouraged to ask for help when needed. All tests were performed in random order.

3. Results

3.1. Research Question 1

Is self-evaluation stable throughout task performance? *Do adolescents exhibit a difference in self-evaluations within a task* (*before*, *during*, *and after a task of reading comprehension and spatial reasoning*)?

In order to examine whether self-evaluation ratings remain stable throughout a task, we ran a one-way repeated measures ANOVA for each of the two tasks separately. Results indicate a significant time effect in the RC task ($F(2, 79) = 41.46$, $p < .001$) and in the TONI task ($F(2, 77) = 10.35$, $p < .001$).

Follow up comparisons indicated that, in both tasks, significant differences were found between confidence ratings before and during the task and confidence ratings after the task (before and during vs. after). More specific, in both tasks, confidence ratings before and during the task were significantly higher

than confidence ratings after the task ($p < .05$). See Table 1.

Next, in order to examine whether confidence ratings (before, during, and after) are one construct for each task, we ran a factor analysis with varimax rotation of the confidence ratings before, during, and after the two tasks. See Table 2.

We found that confidence ratings before, during, and after the RC and TONI tasks loaded on two different constructs, such that one construct includes confidence ratings before, during, and after RC and the other before, during, and after the TONI, with reliability (Cronbach's α) of .86 and .76, respectively.

3.2. Research Question 2

Are academic self-evaluations and calibrations domain-general or task-specific? *Do adolescents exhibit different patterns of self-evaluation and of calibration across tasks* (*reading comprehension and a spatial reasoning task*)?

In order to examine this question, we ran paired-samples t-tests between calibration and confidence ratings before, during, and after the two tasks (reading comprehension and TONI). See Table 3.

The results indicated significant differences in all self-evaluation measures (confidence ratings) and calibration between the RC and TONI tasks, indicating that all confidence ratings across tasks were significantly different from each other, such that confidence ratings before, during, and after the TONI were higher than those confidence ratings in reading comprehension. In additaion, the TONI score and calibration was higher than the RC score and CoC. For descriptive statistics of calibration and confidence ratings in reading comprehension and TONI tasks see Table 1.

Table 1. Descriptive statistics of calibration and confidence ratings before, during, and after reading comprehension and TONI tasks.

	Before	During	After	Calibration	Score
Reading Comprehension	77.03 (13.27)	76.06 (14.21)	66.50 (17.36)	−20.94 (18.18)	55.11 (15.13)
TONI	82.34 (12.96)	84.61 (12.74)	77.32 (17.67)	−6.37 (13.45)	78.24 (11.85)

RC score (range): 13.33 - 86.67; TONI score (range): 51.11 - 97.78; CoC (range): −70.83 - 30; Toni calibration (range): −36.67 - 29.33.

Table 2. Factor analysis pattern matrix.

	Component	
	1	2
Confidence rating before RC task	.677	
Confidence rating during RC task	.935	
Confidence rating after RC task	.882	
Confidence rating before TONI task		.637
Confidence rating during TONI task		.802
Confidence rating after TONI task		.937
% of Variance	**60.87**	**17.34**

Table 3. Paired-samples t-tests between calibration and confidence ratings before, during, and after the two tasks (reading comprehension and TONI).

		Mean	SD	t(1, 84)
pair 1	RC score	−23.12	15.75	−13.61**
	TONI score			
pair 2	Confidence rating before RC task	−4.75	11.98	−3.56**
	Confidence rating before TONI task			
pair 3	Confidence rating during RC task	−8.54	14.39	−5.50**
	Confidence rating during TONI task			
pair 4	Confidence rating after RC task	−12.20	18.34	−5.84**
	Confidence rating after TONI task			
pair 5	CoC	−14.57	16.70	−8.09**
	TONI calibration			

**$p < .005$. RC = Reading Comprehension. CoC = Calibration of Comprehension. According to the Bonferroni correction, the significant differences are $p < .0125$.

Then, we ran Pearson correlations between calibration and confidence ratings before, during, and after the two tasks (reading comprehension and TONI). See **Table 4**.

The results indicated medium to strong correlations among confidence ratings in the two tasks as a factor of time of rating. CoC was moderately associated with TONI calibration (r = .47, $p < .01$).

3.3. Research Question 3

Are self-evaluations and calibration abilities associated with general *reading skill? Do good and poor comprehenders differ in their self-evaluations and calibration scores*?

Constructing the comprehension groups

First, participants were assigned to two distinct reading comprehension groups based on a median split of their scores in the reading comprehension test. The poor comprehenders group (PC) consisted of 43 participants (18 female, 25 male) who scored below the 50th percentile (M = 43.17%, SD = 9.89); the good comprehenders group (GC) consisted of 43 participants (24 female, 19 male) who scored above the 50th percentile (M = 67.05%, SD = 8.58%).

In order to examine whether good and poor comprehenders differ in their confidence ratings and CoC scores, an independent-samples t-test was run between the two groups. Results revealed significant group differences only in CoC, such that good comprehenders were more calibrated than poor comprehenders. **Table 5** shows the descriptive statistics and t-test results for comparisons of the two groups on confidence ratings and CoC.

Bias patterns within comprehension groups

In order to examine more specifically whether confidence rating and calibration

Table 4. Pearson correlations between confidence ratings before, during, and after the reading comprehension and TONI tasks.

		Toni			
		Before	During	After	Calibration
Reading Comprehension	Before	.57**			
	During		.43**		
	After			.37**	
	Calibration				.47**

**$p < .01$.

Table 5. Descriptive statistics and comparisons of confidence ratings and CoC among the two comprehension groups.

	Poor comprehenders (n = 43)		Good comprehenders (n = 43)		t(1, 84)
	M	SD	M	SD	
Confidence rating before RC task	76.16	15.26	77.9	11.03	−.67
Confidence rating during RC task	73.31	16.41	78.81	11.13	−1.82
Confidence rating after RC task	63.42	18.63	69.28	15.83	−1.52
CoC	−30.13	18.01	−11.76	13.1	−5.40**

**$p < .01$. Note. RC = reading comprehension; CoC = calibration of comprehension.

are associated with task or skill we examined the distribution of bias in confidence ratings across comprehension groups (over-confidence vs. under-confidence). Participants with low scores in RC tasks showed the greatest over-confidence in RC, compared to high comprehenders (see **Figure 1**). In addition, not all PC rated relatively low levels of confidence relative to performance, nor did all GC rate relatively high levels of confidence relative to performance (see **Figure 1**).

4. Discussion

In recent years, there has been growing interest in metacognitive self-evaluations as factors influencing reading comprehension (Kasperski & Katzir, 2013; Lin, Moore, & Zabrucky, 2001). Research indicates that students with learning difficulties overestimate their capabilities in academic and non-academic tasks (Job & Klassen, 2012). The current study contributes to the research on academic meta-cognitive evaluation among adolescents. We examined whether these evaluations are domain-general and focused on the differences between poor and good comprehenders.

4.1. Is Self-Evaluation Stable over Performance of a Task?

As expected, results indicate that self-evaluations in RC and in the TONI are not stable and are specific to the ongoing task, regardless of the type of task. We found the same patterns for both tasks, which showed that confidence ratings

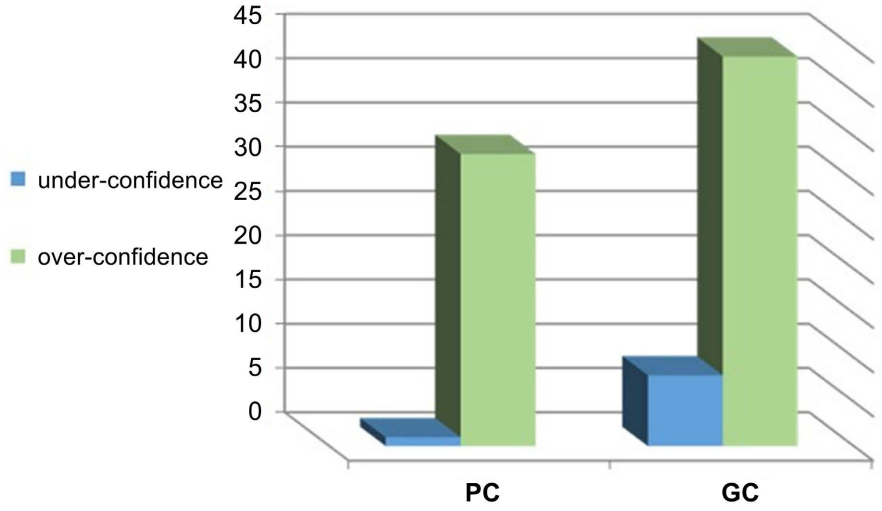

Figure 1. Percentages of students with over and under-confidence ratings in RC, for the two comprehension groups. Note. PC = poor comprehenders group; GC = good comprehenders group.

before and during the task are significantly higher than confidence ratings after the task. Early confidence ratings of the task (before and during the task) may be based on the student's theories or beliefs, for example, beliefs about self-efficacy or about their own ability (Ehrlinger & Dunning, 2003) or expertise (Glenberg & Epstein, 1987). However, these ratings may change after the task since they are based on additional self-feedback generated by taking the test (Devolder, Brigham, & Pressley, 1990; Maki, Jonas, & Kallod, 1994). These findings support the claim that students are able to use additional information generated by performing a test as feedback to make more precise self-evaluations after the task (Lin, Moore, & Zabrucky, 2001). Michalsky, Mevarech, & Haibi (2009) examined 3 treatment groups of 4th grade children who received metacognitive instruction—before reading, during reading, or after reading, with a control group. The findings showed that students who were provided with metacognitive instruction *after* the reading significantly outperformed the other two groups on all outcomes assessed (domain-specific scientific knowledge, scientific literacy, and metacognitive awareness). They conclude that the starting phase includes planning strategies (task analysis and goal setting), the performing phase includes self-control strategies (self-instruction, imagery, attention focusing, task strategies) and self-observation strategies (monitoring, self-recording). The final phase refers to students' strategies for introspecting their learning performance to control and adjust their learning accordingly.

In the current study, factor analysis revealed that confidence ratings before, during, and after the RC and TONI tasks loaded on two different constructs, such that one construct included confidence ratings before, during, and after RC and another for the TONI. We can conclude that, in general, self-evaluations within a task are not consistent and stable, but they are also domain-specific factors since evaluations in both tasks were found to be different and distinct.

Our results confirm that confidence ratings depend on factors that are not stable, such as characteristics of a person but also characteristics of the tasks. Thus, judgments differ within a person across tests.

4.2. Are Academic Self-Evaluations Domain General?

Self-evaluations and calibration across tasks: verbal task vs. spatial reasoning task

As expected, the results showed significant differences *across* tasks (RC vs. TONI) in all self-evaluation measures (confidence ratings) and calibration, such that all confidence ratings in the TONI were higher than in RC, and calibration was better in the TONI as well. This supports the claim that confidence ratings and calibration are mainly domain-specific (Alexander, Dinsmore, Parkinson, & Winters, 2011). Furthermore, the current findings support previous findings, which found that perceptual judgments are more accurate than what people feel, such that participants' responses to perceptual judgments tend to be correct even when the participants feel that their responses were based on a complete guess (Björkman, Juslin, & Winman, 1993). The current study extends previous findings to different complicated tasks, among adolescents.

In addition, as we expected, results showed that confidence ratings before, during, and after the reading comprehension and TONI tasks were moderately associated. Moreover, the results showed that calibration in reading comprehension tasks was moderately associated with calibration of the TONI task. The correlations between confidence ratings from different complex cognitive tests are high enough to define confidence ratings as also domain-general (Kleitman & Stankov, 2001; Nietfeld & Schraw, 2002; Pallier et al., 2002; Stankov & Crawford, 1996, 1997).

While most of the studies focused on young children and adults only (Ackerman & Leiser, 2014; Glover, 1989; Klassen, 2007; Kasperski & Katzir, 2013) and examine self-evaluation in basic cognitive tasks or comparisons of tasks from the same domain, our study added knowledge about high school aged adolescents and compared two complex tasks from different domains.

4.3. Are Confidence Ratings and Calibration Associated with Reading Comprehension Skills?

Differences between the comprehension groups

Our results indicated that good comprehenders were better than poor comprehenders in CoC. These findings suggest that, among adolescents, higher levels of comprehension are related to higher levels of metacognitive skill. Overall, no differences emerged between the two comprehension groups in RC confidence ratings before, during, and after the task. Differences were found only in calibration. Therefore, poor comprehenders have the same confidence ratings as good comprehenders, but due to their lower scores in the reading comprehension task, the difference between their confidence ratings and performance (miscalibration) is the greatest. These findings support previous findings of over-

confidence in adults and children with reading disabilities (Kleitman & Stankov, 2001, 2007; Maki et al., 2005; Nietfeld et al., 2005; Kasperski & Katzir, 2013; West & Stanovich, 1997; Winne & Jamieson-Noel, 2002) and extend them to adolescents.

Results further suggest that poor comprehenders may have a "double curse" (Dunning, Johnson, Ehrlinger, & Kruger, 2003): a reading challenge and a failure to recognize it. It is possible that the same monitoring problems that impact their reading comprehension evaluations underlie their inability to accurately evaluate their performance. This finding suggests that poor comprehenders not only possess poor general knowledge about useful reading comprehension strategies (de Carvalho Filho & Yuzawa, 2001; Kleitman & Stankov, 2007), but also lack knowledge about their performance in a particular reading comprehension situation.

It is also possible that students with low comprehension may overestimate their true capabilities as a means of ego protection. For example, students with LD tend to explain the overestimations of other students in terms of ego-protective motivations (e.g., "They didn't want to admit they didn't know") while simultaneously explaining their own overestimations in terms of misjudgment ("I really thought I could do the problems") (Stone & May, 2002).

Finally, our bias analysis found that the distribution of confidence levels across the comprehension groups was supportive of the trait-driven approach (Kleitman & Stankov, 2001; Stankov & Crawford, 1996, 1997). Thus, even among the highest performing students there were those who were insecure about their performance, and among the lowest-performing students there were those who felt secure about their performance. In sum, the distribution of the confidence judgments indicates that confidence ratings among high school students are affected not only by reading comprehension ability but also by personality factors or traits.

In conclusion, our findings suggest that metacognitive evaluations are domain-specific as well as domain-general. Specifically, self-evaluation and calibration in RC were different from those measured in the TONI and these evaluations loaded on two different constructs. However, when we looked within each task we found the same patterns for both tasks, which showed that confidence ratings before and during the task were significantly higher than confidence ratings after the task. Moreover, calibration is also influenced by reading comprehension skills (Lin, Moore, & Zabrucky, 2001; Lin, Zabrucky, & Moore, 1997). GC groups were better than PC groups in CoC. Further, poor comprehenders have the same confidence ratings as other groups, but their performance is lower. Thus, the difference between their confidence ratings and their performance (overconfidence) is the greatest. Adolescents who are poor comprehenders display a "double curse," i.e., both a reading challenge and a failure to recognize it. In terms of their self-regulation measures by CoC they were indeed lower. These findings strongly suggest that work with adolescents should focus not only on

strategies but also on self-regulation. In addition, our findings suggest that metacognitive evaluations are domain-specific as well as domain-general. There is an individual tendency to be confident or less confident across content areas, but it is also driven by skill.

4.4. Implications

The results of this study have several meaningful educational implications. Adolescents, similar to children and adults, are not always calibrated in their reading comprehension evaluations compared to other domains such as non-verbal tasks. Thus, introducing a meta-cognitive approach to reading from an early stage, when students enter the educational system, will prove important for metacognitive self-evaluation judgments in learning when reading texts.

4.5. Limitations of the Study and Future Directions

This study has several limitations that concern the socioeconomic background of our sample, as well as the sample size. That is, although our sample was relatively heterogeneous, both schools were in the center of the country and had a similar socioeconomic status. Therefore, future studies should determine whether the results are replicable in a more socioeconomically diverse population and with a larger sample. Moreover, this study was conducted with adolescents in tenth grade. In the future, longitudinal and cross-sectional studies should examine the developmental aspects of self-evaluations across domains. In addition, the current study focused on a comparison of two cognitive tasks. It would be interesting to examine whether the findings would be similar when comparing academic and non-academic tasks, such as motor, graphic, emotional tasks, etc. Finally, we examined whether calibration is domain-general; it would be interesting to examine which specific personality traits relate to this ability, such as self-awareness, the tendency to attribute failures to internal causes, and self-esteem.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

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,* 1816-1828. https://doi.org/10.1016/j.chb.2012.04.023

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,* 329-348. https://doi.org/10.1111/bjep.12021

Alexander, P. A., Dinsmore, D. L., Parkinson, M. M., & Winters, F. I. (2011). Self-Regulated Learning in Academic Domains. In Dale H. Schunk, Barry Zimmerman (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 393-407). Routledge.

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, 2,* 251-276. https://doi.org/10.2307/747763

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval Fluency as a Metacognitive Index. In Lynne M. Reder (Ed.), *Implicit Memory and Metacognition* (pp. 309-338). Psychology Press.

Björkman, M., Juslin, P., & Winman, A. (1993). Realism of Confidence in Sensory Discrimination: The Underconfidence Phenomenon. *Attention, Perception, & Psychophysics, 54,* 75-81. https://doi.org/10.3758/BF03206939

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Nonverbal Intelligence: TONI-4.* Pro-ed.

Bryant, B. R., & Wiederholt, J. L. (1991). *Gray Oral Reading Test-Diagnostic*. Pro-Ed.

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,* 372-386. https://doi.org/10.1016/j.lindif.2008.10.004

de Carvalho Filho, M. K., & Yuzawa, M. (2001). The Effects of Social Cues on Confidence Judgments Mediated by Knowledge and Regulation of Cognition. *The Journal of Experimental Education, 69,* 325-343. https://doi.org/10.1080/00220970109599491

Devolder, P. A., Brigham, M. C., & Pressley, M. (1990). Memory Performance Awareness in Younger and Older Adults. *Psychology and Aging, 5,* 291-303. https://doi.org/10.1037/0882-7974.5.2.291

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. https://doi.org/10.1016/j.learninstruc.2012.06.001

Duke, N. K., & Pearson, P. D. (2009). Effective Practices for Developing Reading Comprehension. *Journal of Education, 189,* 107-122. https://doi.org/10.1177/0022057409189001-208

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why People Fail to Recognize Their Own Incompetence. *Current Directions in Psychological Science, 12,* 83-87. https://doi.org/10.1111/1467-8721.01235

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. https://doi.org/10.1023/A:1007996502372

Ehrlinger, J., & Dunning, D. (2003). How Chronic Self-Views Influence (and Potentially Mislead) Estimates of Performance. *Journal of Personality and Social Psychology, 84,* 5-17. https://doi.org/10.1037/0022-3514.84.1.5

Falchikov, N., & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research, 59,* 395-430. https://doi.org/10.3102/00346543059004395

Finney, S. J., & Schraw, G. (2003). Self-Efficacy Beliefs in College Statistics Courses. *Contemporary Educational Psychology, 28,* 161-186. https://doi.org/10.1016/S0361-476X(02)00015-2

Flavell, J. H. (1976). Metacognitive Aspects of Problem Solving. In L. B. Resnick (Ed.), *The Nature of Intelligence* (pp. 231-235). Lawrence Erlbaum Associate.

Garcia, T., & Pintrich, P. R. (1994). Regulating Motivation and Cognition in the Classroom: The Role of Self-Schemas and Self-Regulatory Strategies. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Self-Regulation of Learning and Performance: Issues and Educa-*

tional Applications (pp. 127-153). Lawrence Erlbaum Associates, Inc.

Glenberg, A. M., & Epstein, W. (1985). Calibration of Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 702-718. https://doi.org/10.1037/0278-7393.11.1-4.702

Glenberg, A. M., & Epstein, W. (1987). Inexpert Calibration of Comprehension. *Memory & Cognition, 15,* 84-93. https://doi.org/10.3758/BF03197714

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing Calibration of Comprehension. *Journal of Experimental Psychology: General, 116,* 119-136. https://doi.org/10.1037/0096-3445.116.2.119

Glover, J. A. (1989). Improving Readers' Estimates of Learning from Text: The Role of Inserted Questions. *Literacy Research and Instruction, 28,* 68-75. https://doi.org/10.1080/19388078909557976

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and Cognitive Predictors of Text Comprehension and Reading Amount. *Scientific Studies of Reading, 3,* 231-256. https://doi.org/10.1207/s1532799xssr0303_3

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test Prediction and Performance in a Classroom Context. *Journal of Educational Psychology, 92,* 160-170. https://doi.org/10.1037/0022-0663.92.1.160

Hadwin, A. F., & Webster, E. A. (2013). Calibration in Goal Setting: Examining the Nature of Judgments of Confidence. *Learning and Instruction, 24,* 37-47. https://doi.org/10.1016/j.learninstruc.2012.10.001

Harter, S. (1998). The Development of Self-Representations. In W. Damon, & N. Eisenberg (Eds.), *Handbook of Child Psychology: Social, Emotional, & Personality Development* (5th ed., Vol. 3, pp. 553-618). Wiley.

Job, J. M., & Klassen, R. M. (2012). Predicting Performance on Academic and Non-Academic Tasks: A Comparison of Adolescents with and without Learning Disabilities. *Contemporary Educational Psychology, 37,* 162-169. https://doi.org/10.1016/j.cedpsych.2011.05.001

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, 34,* 59-84. https://doi.org/10.1080/02702711.2011.580042

Klassen, R. M. (2007). Using Predictions to Learn about the Self-Efficacy of Early Adolescents with and without Learning Disabilities. *Contemporary Educational Psychology, 32,* 173-187. https://doi.org/10.1016/j.cedpsych.2006.10.001

Kleitman, S., & Stankov, L. (2001). Ecological and Person-Oriented Aspects of Metacognitive Processes in Test-Taking. *Applied Cognitive Psychology, 15,* 321-341. https://doi.org/10.1002/acp.705

Kleitman, S., & Stankov, L. (2007). Self-Confidence and Metacognitive Processes. *Learning and Individual* Differences, 17, 161-173. https://doi.org/10.1016/j.lindif.2007.03.004

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014). The Effects of Goal-Driven and Data-Driven Regulation on Metacognitive Monitoring during Learning: A Developmental Perspective. *Journal of Experimental Psychology: General, 143,* 386-403. https://doi.org/10.1037/a0031768

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The Memorizing Effort Heuristic in Judgments of Learning: A Developmental Perspective. *Journal of Experimental Child Psychology, 102,* 265-279. https://doi.org/10.1016/j.jecp.2008.10.005

Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Per-*

sonality and Social Psychology, 77, 1121-1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lenz, B. K., Ellis, E., & Scanlon, D. (1996). *Teaching Learning Strategies to Adolescents and Adults with Learning Disabilities*. Pro-Ed.

Lin, L. M., & Zabrucky, K. M. (1998). Calibration of Comprehension: Research and Implications for Education and Instruction. *Contemporary Educational Psychology, 23,* 345-391. https://doi.org/10.1006/ceps.1998.0972

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,* 111-128. https://doi.org/10.1080/027027101300213083

Lin, L., Zabrucky, K., & Moore, D. (1997). The Relations among Interest, Self-Assessed Comprehension, and Comprehension Performance in Young Adults. *Reading Research and Instruction, 36,* 127-139. https://doi.org/10.1080/19388079709558233

Maki, R. H. (1995). Accuracy of Metacomprehension Judgments for Questions of Varying Importance Levels. *The American Journal of Psychology, 108,* 327-344. https://doi.org/10.2307/1422893

Maki, R. H. (1998). Test Predictions over Text Material. *Metacognition in Educational Theory and Practice, 14,* 117-144.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of Text Material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 663-679. https://doi.org/10.1037/0278-7393.10.4.663

Maki, R. H., & Serra, M. (1992). The Basis of Test Predictions for Text Material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 116-126. https://doi.org/10.1037/0278-7393.18.1.116

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased Processing Enhances Calibration of Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 609-616. https://doi.org/10.1037/0278-7393.16.4.609

Maki, R. H., Jonas, D., & Kallod, M. (1994). The Relationship between Comprehension and Metacomprehension Ability. *Psychonomic Bulletin & Review, 1,* 126-129. https://doi.org/10.3758/BF03200769

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual Differences in Absolute and Relative Metacomprehension Accuracy. *Journal of Educational Psychology, 97,* 723-731. https://doi.org/10.1037/0022-0663.97.4.723

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,* 99-108. https://doi.org/10.1111/j.1540-5826.2004.00093.x

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

Nietfeld, J. L., & Schraw, G. (2002). The Effect of Knowledge and Strategy Training on Monitoring Accuracy. *The Journal of Educational Research, 95,* 131-142. https://doi.org/10.1080/00220670209596583

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom. *The Journal of Experimental Educational, 74,* 7-28.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L. et al. (2002).

The Role of Individual Differences in the Accuracy of Confidence Judgments. *Journal of General Psychology, 129,* 257-299. https://doi.org/10.1080/00221300209602099

Pieschl, S. (2009). Metacognitive Calibration—An Extended Conceptualization and Potential Applications. *Metacognition and Learning, 4,* 3-31. https://doi.org/10.1007/s11409-008-9030-4

Pintrich, P. R. (1988). A Process-Oriented View of Student Motivation and Cognition. *New Directions for Institutional Research, 57,* 65-79. https://doi.org/10.1002/ir.37019885707

Pressley, M., & Ghatala, E. S. (1989). Metacognitive Benefits of Taking a Test for Children and Young Adolescents. *Journal of Experimental Child Psychology, 47,* 430-450. https://doi.org/10.1016/0022-0965(89)90023-4

Pressley, M., Levin, J. R., Ghatala, E. S., & Ahmad, M. (1987). Test Monitoring in Young Grade School Children. *Journal of Experimental Child Psychology, 43,* 96-111. https://doi.org/10.1016/0022-0965(87)90053-1

Prior, A., Zeltsman-Kulick, R., & Katzir, T. (2020). Adolescent Word Reading in English as a Foreign Language. *Journal of Research in Reading, 43,* 116-139. https://doi.org/10.1111/1467-9817.12293

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,* 31-46. https://doi.org/10.1207/S15327752JPA7801_03

Ruban, L. M., McCoach, D. B., McGuire, J. M., & Reis, S. M. (2003). The Differential Impact of Academic Self-Regulatory Methods on Academic Achievement among University Students with and without Learning Disabilities. *Journal of Learning Disabilities, 36,* 270-286. https://doi.org/10.1177/002221940303600306

Schiefele, U. (1991). Interest, Learning, and Motivation. *Educational Psychologist, 26,* 299-323. https://doi.org/10.1080/00461520.1991.9653136

Schraw, G. (2009). A Conceptual Analysis of Five Measures of Metacognitive Monitoring. *Metacognition and Learning, 4,* 33-45. https://doi.org/10.1007/s11409-008-9031-3

Schraw, G., & Roedel, T. D. (1994). Test Difficulty and Judgment Bias. *Memory & Cognition, 22,* 63-69. https://doi.org/10.3758/BF03202762

Stankov, L., & Crawford, J. D. (1996). Confidence Judgments in Studies of Individual Differences. *Personality and Individual Differences, 21,* 971-986. https://doi.org/10.1016/S0191-8869(96)00130-4

Stankov, L., & Crawford, J. D. (1997). Confidence and Performance on Tests of Cognitive Abilities. *Intelligence, 25,* 93-109. https://doi.org/10.1016/S0160-2896(97)90047-7

Stipek, D. (1998). *Motivation to Learn*. Allyn and Bacon.

Stolp, S., & Zabrucky, K. M. (2009). Contributions of Metacognitive and Self-Regulated Learning Theories to Investigations of Calibration of Comprehension. *International Electronic Journal of Elementary Education, 2,* 7-31.

Stone, C. A., & May, A. L. (2002). The Accuracy of Academic Self-Evaluations in Adolescents with Learning Disabilities. *Journal of Learning Disabilities, 35,* 370-383. https://doi.org/10.1177/00222194020350040801

Thiede, K. W., Anderson, M., & Therriault, D. (2003). Accuracy of Metacognitive Monitoring Affects Learning of Texts. *Journal of Educational Psychology, 95,* 66-73. https://doi.org/10.1037/0022-0663.95.1.66

Vadhan, V., & Stander, P. (1994). Metacognitive Ability and Test Performance among College Students. *The Journal of Psychology, 128,* 307-309.

https://doi.org/10.1080/00223980.1994.9712733

Weaver, C. (1990). *Understanding Whole Language*. Heinemann Educational Books.

West, R. F., & Stanovich, K. E. (1997). The Domain Specificity and Generality of Overconfidence: Individual Differences in Performance Estimation Bias. *Psychonomic Bulletin & Review, 4,* 387-392. https://doi.org/10.3758/BF03210798

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). The Guilford Press.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the Comprehension in Metacomprehension. *The Journal of General Psychology, 132,* 408-428. https://doi.org/10.3200/GENP.132.4.408-428

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring Students' Calibration of Self Reports about Study Tactics and Achievement. *Contemporary Educational Psychology, 27,* 551-572. https://doi.org/10.1016/S0361-476X(02)00006-1

Abbreviations

RC: Reading comprehension

CoC: Calibration of comprehension

PC: Poor comprehenders

GC: Good comprehenders