Metacognition and Learning (2023) 18:113–134
https://doi.org/10.1007/s11409-022-09317-8

Metacognitive monitoring skills of reading comprehension and writing between proficient and poor readers

Christian Soto[1] · Antonio P. Gutierrez de Blume[2] · Verónica Rebolledo[1] · Fernanda Rodríguez[1] · Diego Palma[1] · Fernando Gutiérrez[1]

Received: 21 January 2022 / Accepted: 22 July 2022 / Published online: 31 August 2022

Abstract
Reading comprehension and writing are essential skills for success in modern societies. Additionally, reading and writing have been described as highly reflective activities that necessitate metacognitive monitoring and control. However, reading comprehension and writing are skills moderated by many factors, proficiency among them. Thus, in the present study we examined the influence of reading comprehension proficiency (proficient, poor) on elementary school students' ($N=120$) metacognitive monitoring accuracy in reading and writing tasks. Further, we investigated the predictive patterns of linguistic indices between proficient and poor readers on their metacognitive monitoring accuracy in a writing task. Findings revealed that proficient readers exhibited significantly better monitoring accuracy in both reading and writing tasks, and that unique predictive patterns of linguistic indices on writing skill monitoring accuracy emerged between proficient and poor readers. We discuss the implications of these findings for research, theory, and practice and propose recommendations for future research.

Keywords Calibration accuracy · Mental representation · Metacognition · Reading proficiency · Writing production

Introduction

Reading and writing are transversal skills to all areas of learning. They are involved in academic and work performance, allow individuals to overcome vulnerable contexts, establish better relationships with the environment, and contribute to social inclusion (UNESCO, 2015). In the 21st century, a large part of human activities involves literacy skills, which

✉ Antonio P. Gutierrez de Blume
agutierrez@georgiasouthern.edu

[1] Department of Spanish, Psycholinguistics Area, Universidad de Concepción, Victor Lamas 1290, Zip Code 4070386, Concepción, Chile

[2] Department of Curriculum, Foundations, and Reading, Georgia Southern University, P.O. Box 8144, 30460-8144 Statesboro, GA, USA

is why they become crucial to achieve success in modern occupations and in social communication (Jacovina & McNamara, 2017). From theory, reading comprehension assumes an active process of meaning construction from textual information and the reader's knowledge (Kintsch, 1998). The outcome of this process is a memory representation that selects certain kinds of information as important and subsumes details into more generalized statements (Kintsch, 1990). On the other hand, writing involves externalizing a series of internal mental representations through prose (Flower & Hayes, 1981). As can be seen, the quality of the mental representation plays a fundamental role in both processes.

Both reading and writing involve low-level instrumental processes (e.g., decoding or transcription of the text) and processes of higher semantic complexity linked to linguistic comprehension in reading or coherent and cohesive writing (for more details, review the models *the simple view of reading* (Hoover & Gough, 1990) and *the simple view of writing* (Berninger, 2000)). This research will be focused on high-level processes and the components that regulate them.

There are three perspectives associated with the intertwined relation between reading and writing: (1) reading and writing are separable processes, with independent operations and representations and without mutual influence; (2) both are separable processes, with some shared representations and operations that are brought together under certain circumstances; and (3) both are inseparable processes, with shared and indivisible operations and representations (Meyer et al., 2016). Although there is no empirical evidence demonstrating that reading and writing are separable and independent processes, there is strong evidence showing the confluence of reading comprehension and writing production (Buz et al., 2016; Fricke et al., 2016; Guzzardo Tamargo et al., 2016; Hsiao & MacDonald, 2016; Kittredge & Dell, 2016; Zamuner et al., 2016).

In addition, Fitzgerald and Shanahan (2000) describe that reading and writing requires meta-knowledge. This meta-knowledge is related to several subcategories of knowledge, highlighting the monitoring executed by readers and writers regarding their ability to generate and create meaning, monitor the identification of words, or the use of strategies to understand or produce and monitor one's own knowledge.

The concept of shared processes and representations in reading and writing has led to the proliferation of studies that have attempted to demonstrate the reciprocal influence of one process on the other (Graham & Hebert, 2011; Graham et al., 2018; Gutierrez de Blume et al., 2021). For example, one of our previous studies aimed to investigate the effect of reading comprehension level (proficient, poor) on students' general writing performance on specific narrative and informative writing tasks ($N=105$). Results revealed that proficient readers outperformed poor readers on objective measures of text production and informative/expository texts. Additionally, regression models demonstrated that proficient readers relied more on deeper aspects of reading and writing such as inferential skills, whereas poor readers tended to focus on superficial aspects of texts, or what Kintsch referred to as text-base, and appeared to perform better in reading and writing tasks related to narratives compared to information-based, expository texts (Gutierrez de Blume et al., 2021). Based on this evidence, we propose an iterative cycle between reading and writing, where the level of coherence obtained during the reading process affects the mental representation of the text, and this mental representation influences the results of reading comprehension and, consequently, of writing. However, the study did not incorporate metacognitive variables, an aspect that we wish to deepen in the present investigation.

Contributors to performance in Reading and writing

It is well established that many children and adolescents have inadequate reading comprehension or writing skills (NAEP, 2003; OECD, 2017). Several studies have attempted to explain the differences between individuals with proficient or poor performance in reading and writing, and the variables that influence these performances (Beauvais et al., 2011; Cain et al., 2004; Ferrari et al., 1998; Honeycutt, 2002; Logan et al., 2011).

In reading, according to the construction-integration (C-I) model (Kintsch, 1988, 1998), deep understanding of the text is achieved by generating a mental representation from the text and the reader's prior knowledge during this process. Reading and comprehension tasks are considered to involve a dynamic interaction between the reader, a text and its context. Riffo et al., (2013) propose three criteria present during the reading process, namely: (1) criteria determined by the level of processing required by the task, considered "textual" comprehension; (2) criteria determined by the context, also called "pragmatic" comprehension; and (3) criteria determined by the reader and his/her position against the text and its contexts, called "critical" comprehension. Perfetti et al., (2005) proposed that in the early stages of learning to read, oral language comprehension, lexical knowledge, and decoding are key elements to achieve comprehension of a text. Later, with more experience and reading practice, students reach higher levels of comprehension where inferential processes and metacomprehension monitoring take on greater relevance. These allow the reader to apply higher standards of coherence, forming representational levels of greater complexity from what has been read. Many of these skills are acquired through implicit learning but can also be improved through explicit teaching or the use of comprehension strategies (Lika, 2017; McNamara, 2007).

In writing, performance is usually determined by an analysis of the written production of individuals. This analysis can be done through the evaluation of mechanical aspects of writing such as spelling, punctuation, grammar, and sentence structure. However, to evaluate the adequate communication of ideas through writing, the evaluation of coherence and textual cohesion become relevant (Struthers et al., 2013). Nevertheless, the automatic evaluation of texts (in terms of coherence and cohesion) is a complex challenge and has been addressed by different researchers (Crossley et al., 2019; Palma et al., 2021; Smith et al., 2016). Initially, formulas such as the Flesch Reading Ease Readability Formula, Flesch-Kincaid Grade Level, New Dale-Chall Readability Formula, among others, were used for the evaluation of readability, but they have been criticized. The main criticism is that these formulas only measure some factors of textual surface, but not the actual sources of difficulty in a text (Palma & Soto, 2022). Fortunately, thanks to advances in computational linguistics, it has been possible to develop computational tools that are able to measure readability based on various linguistic indices (e.g., sentence length, clause length, lexical diversity, etc.), that consider several levels of discourse (e.g., surface proxies, situational model, genre and rhetorical structure). Some of these tools are Coh-metrix (Graesser et al., 2004) and TRUNAJOD (Palma et al., 2019, 2021). The value of such tools has been recognized, as the evaluation of the complexity of the text is a natural language processing task that can be extended to multiple applications such as the automatic evaluation of abstracts, automatic evaluation of essays, evaluation of open questions, intelligent tutoring systems, etc. Writing quality is highly related to the cohesion and coherence of the prose; the more cohesion and coherence, the higher the quality of the writing and vice versa (Grabe & Kaplan,

2014; Spiegel & Fitzgerald, 1990; Witte & Faigley, 1981; Yang & Sun, 2012). In addition, various scholars view coherence as partly residing in the reader's mind. In that perspective, coherence does not arise solely from textual factors, but also from the reader's own schemata (Gernsbacher & Givón, 1995; Lee, 2004). From that context, Hayes (2012) analyzed a corpus of written production of first to ninth-grade students and identified three kinds of knowledge-telling: (1) flexible-focus texts, which do not have a central topic, but rather they can be a sequence of statements about the immediately previous text; (2) fixed-topic texts, where each utterance in the text refers to the same topic; and (3) topic-elaboration texts, which have an organized structure with a global topic and subtopics. Also, the first two types of texts are most common in the early grades, and, after the sixth grade, the topic-elaboration texts are most common. In relation to performance, proficient writers transform knowledge when writing, whereas poor writers often simply engage in "knowledge telling" without attempting to clearly organize their ideas (Bereiter & Scardamalia, 1987).

In addition to linguistic, cognitive, and motivational variables, the reading and writing processes have been described as highly reflective activities that necessitate metacognitive monitoring and control (Bereiter & Scardamalia, 1987). Thus, several studies have explored the influence of metacognition (Dunlosky & Lipko, 2007; Hacker et al., 1998; Wiley et al., 2005; Wong, 1999). These studies concluded that, both in reading and writing, metacognitive abilities seem to explain the differences in the performance between skilled readers and writers and less-skilled ones.

Role of Metacognition in Reading and writing

Metacognition is the knowledge and cognition about cognitive phenomena (Flavell, 1978, 1979). It is based on two key components that are closely related, knowledge about cognition or metacognitive awareness, and regulation of cognition. Both metacognitive components are important to enhance performance in cognitive tasks (Soto et al., 2018) and to transfer the learned knowledge to new situations (Halpern, 1998). However, due to the flexibility of its application, the focus of this study will be on metacognitive regulation.

Readers who successfully monitor and regulate their comprehension know when they understand, when they do not understand, and when they partially understand. Also, they know how to evaluate if their comprehension is adequate for their reading goals and how to deal with comprehension difficulties (Baker, 1979). From these findings, it is possible to conceive metacomprehension regulation as an essential skill for competent reading because it directs the reader's cognitive processes to make sense of incoming textual information (Westby, 2004). Comparative studies between proficient and poor readers confirm this assertion, given that in several studies low-performing readers tend not to monitor their understanding (Baker, 1984; Ehrlich, 1996; Ehrlich et al., 1999).

In writing, translating ideas into text might lead one to think of new ideas or to revise an earlier part of one's text to achieve consistency, before returning to continue with the translation process. To successfully conduct such evaluations and adjustments, self-regulation of the writing process is essential (Ferrari et al., 1998). Research suggests that, as compared to poor writers, skilled writers are more actively and more metacognitively involved in the writing process; that is, skilled writers spend more time recursively planning and revising

their text (Harris & Graham, 1992; Hayes & Flower, 1980) and they are more active at monitoring their writing (Beal et al., 1990).

To evaluate the levels of regulation and metacognitive monitoring, one of the measures commonly used is *absolute* monitoring accuracy (henceforth called "calibration accuracy"). This refers to the accuracy of students' perceptions of their own performance on a cognitive task, in this case, reading or writing. Accurate calibration implies a high alignment between the judgment of performance on the task and actual performance. On the contrary, inaccurate calibration implies an inconsistency between the judgments about what individuals know and do not know about the task and actual performance. In this case, there may be a tendency towards overconfidence in performance or towards lack of confidence (Pintrich et al., 2000), both of which are manifestations of monitoring error. Calibration accuracy has a fundamental role in learning because it allows individuals to self-regulate their effort, promoting self-regulated learning (Stone, 2000). Research also has shown that calibration accuracy increases with scaffolding techniques; it is positively related to prior knowledge and to students' performance; it improves when the students are given time to study and when the judgments of comprehension are delayed; and when students receive feedback and encouragement (Gutierrez de Blume, 2020; Gutierrez & Price 2017; Gutierrez & Schraw, 2015). Also, research has demonstrated the combined effect of incentives and training on performance, confidence, and calibration accuracy, showing that strategy training combined with incentives has the greatest learning effect on adults (Gutierrez & Schraw, 2015; Gutierrez de Blume, 2020) and on 9- and 10-year-old children (Gutierrez de Blume, 2017).

Despite the large body of research in the field of literacy and metacognition, the relationships between reading and writing performance with calibration accuracy, as a measure of metacognitive regulation, have yet to be explored. It would be interesting to determine if in writing there are elements of written production that function as metacognitive cues with influence on the process itself, as it happens in reading (Soto et al., 2020), serving as a variable to differentiate poor and proficient writers. It is worth mentioning that in this research the terms "proficient" and "poor" readers will be used specifically in relation to performance in reading comprehension of texts.

Link between Metacognitive Monitoring and Reading Comprehension

The relation between students' confidence in performance judgments and their actual performance is considered a measure of metacognitive monitoring skill, of which calibration accuracy and bias (metacognitive errors) are indices. When students successfully self-rate their level of reading comprehension, they should be commensurately accurate in their confidence in performance judgments (either item-by-item [locally] or globally before [prediction] and/or after [postdiction] an assessment). Nevertheless, recent research on the topic suggests that students' metacognitive monitoring skills in reading are lacking (Gutierrez de Blume et al., 2021; Soto et al., 2018). More specifically, a series of studies tasked students to read a text, make a confidence in performance judgment about how they think they will perform on a reading comprehension test, and then complete a reading performance test (e.g., Bol & Hacker 2001; Bol et al., 2005; Dunlosky & Lipko, 2007; Hacker et al., 2008). Even though many of the confidence in performance judgments are regarding students' overall comprehension level (e.g., asking how well students think they will do on an upcom-

ing test by asking, for example, item-by-item whether they think they obtained a correct response on multiple-choice items), even when students are asked to predict performance regarding specific bits of information from a text (e.g., asking how well students think they will be able to recall a definition [global judgment]), reading metacomprehension accuracy is still quite low and, conversely, errors in reading metacomprehension abound (e.g., incorrect reading performance that is judged to be correct, or an *illusion of knowing*, or correct reading performance judged to be incorrect, or an *illusion of not knowing*; Dunlosky et al., 2002; Gutierrez de Blume et al., 2021; Soto et al., 2018).

Thiede and colleagues (2009) conducted an extensive analysis of the relations between metacognitive monitoring and reading comprehension. They found that the average correlation in more than 40 studies was approximately, $r = .27$. Thiede et al. attributed this relatively low correlation by describing several concerns that could undermine students' confidence in performance judgments relative to their actual reading comprehension performance, or that could lead to low levels of accuracy in metacognitive monitoring judgments (e.g., overconfident students who judge incorrect performance as correct or underconfident students who judge correct performance as incorrect). The authors argue that a lack of evidence supporting the validity of scores from the reading comprehension tests used in previous studies exists. Weaver (1990), for instance, demonstrated that there are improvements to reading metacomprehension accuracy when several elements of the text (e.g., text base [superficial] or model base [deep, as it requires drawing inferences]) are included in the reading comprehension test. In addition, the test should incorporate information from several elements of the text (e.g., explicit and implicit, vocabulary knowledge, relation between paragraphs or ideas, global comprehension, etc.; Dunlosky et al., 2005; Soto et al., 2018). Therefore, evidence implied that measures of reading comprehension should include various components of the text and the relations among these components. When a measure of reading comprehension focuses too heavily on certain material, the association between confidence in performance judgments and actual performance will, in general, decrease. Dunlosky and Lipko (2007) also noted that reading metacomprehension accuracy can be impacted by text length, such that for longer texts readers will have more difficulty producing accurate confidence in performance judgments about their actual reading comprehension. Similarly, prior research indicated that the accuracy of metacomprehension is greater when the text's reading level is approximating the student's actual reading skill level (Soto et al., 2018; Weaver & Bryant, 1995).

Along a similar vein, additional research posits that poor reading metacomprehension accuracy results when readers employ inappropriate text cues to generate their confidence in performance judgments. Dunlosky et al. (2002), for example, proposed the *levels of disruption* theory, which argues that when readers make confidence in performance judgments about a text, they base these judgments on cues that are derived from three possible dilemmas that undermine their comprehension of text: (1) *inference assumption*; (2) *accuracy assumption*; and (3) *representation assumption.* The *representation assumption*, for instance, stipulates that text comprehension difficulties can occur at different levels of representation of the text. When difficulties occur primarily at one specific level, readers' metacomprehension is based on that level of the text. Depending on readers' comprehension level, they employ various cues to judge their level of comprehension. Because they operate at a particular level of comprehension, the reading comprehension difficulty will occur in different elements of this representation. More specifically, if novice readers operate in their

reading process by generating basic relations between explicit information, the dissonance between explicit textual inconsistencies will result in a dissonance in their representation, affecting the fluency of their reading. Conversely, if more advanced readers operate in a situation model dimension, in which they generate deep connections between the information of the text and their prior knowledge, the reading difficulty will occur at a more advanced level, affecting the fluency of their reading comprehension differently when compared to novice readers. Thus, it is plausible that some information that creates reading difficulties for novice readers will not affect advanced readers and vice-versa.

In fact, this relation between accuracy, knowledge, and reading comprehension skills is supported by extant research.

Thiede et al., (2010) investigated the association between a metacognitive strategy (i.e., concept map) and the understanding of texts among university students. They found that students' metacomprehension monitoring in metacognitive strategies around information from a situation model level was reflected in their reading comprehension performance. Maki et al. (2005) investigated the absolute and relative accuracy of reading metacomprehension. After reading unknown and revised texts, students answered a multiple-choice test. Results revealed that the known texts produced an overconfidence (i.e., incorrect performance judged to be correct) in the readers and greater reading metacomprehension errors than the texts that seemed unknown to the students. Likewise, Maki and Berry (1984) suggested that readers, when examining a text, generate at least two types of judgments: judgments about their comprehension of the text and judgments about knowledge of the text's content. Indeed, the relation between reading and metacognitive monitoring could affect not only reading comprehension (Soto et al., 2018) but also other dimensions of reading. Similarly, Chen et al. (2009) examined the reading awareness of college students by reading expository texts. Findings indicated that phonological awareness, phonetics, reading fluency, vocabulary development and reading comprehension are, in some way, affected by metacognitive monitoring skill level.

The Present Study

It would be innovative to know if there are relationships between reading and writing calibration accuracy, considering a poor and proficient performance. Developing a better understanding of the differences between proficient and poor readers and writers and their relation to metacognitive dimensions is essential because such findings could enrich the instructional models on reading and writing, considering metacognition as the underlying mechanism that explains these differences. With this in mind, the present study was guided by the following research questions and expectations.

Research Questions and Hypotheses

1. Are there differences between proficient and poor readers in metacognitive monitoring accuracy (prediction and postdiction monitoring judgments) in reading and writing tasks?

.Hypothesis 1 We expected that proficient readers will have significantly better metacognitive monitoring accuracy (prediction and postdiction) in reading and in writing when compared to poor readers.

2. What are the differences in the predictive patterns of linguistic indices (extracted from the writing task) on writing monitoring accuracy between proficient and poor readers?

Hypothesis 2 We expected that there would be distinct predictive patterns regarding linguistic indices for proficient and poor readers, as shown in previous research (Soto et al., 2019; Soto et al., 2020).

Method

Participants, Research Design, and Sampling

The present study was based on a convenience sampling procedure to recruit participants and a non-experimental correlational research design that employed a combination of descriptive and inferential statistics.

The participants were 120 Chilean students (54 girls and 66 boys) in fifth grade. None of the participants was diagnosed with a reading or writing learning disability as of the completion of this study. The students belonged to three public schools in the community of Concepción. The schools were located in the same geographic area, and the students of these three schools had a similar socioeconomic level (according to criteria of the Ministry of Education). In addition, the schools selected in this study have shown appropriate educational results, according to the verifications carried out by the Ministry of Education. The native language of the participants is Spanish. The average age of the participants was 10 years and 6 months.

Instruments and Materials

Reading comprehension proficiency

Reading comprehension was measured using *Lectum Level 4*. *Lectum* is a reading comprehension test initially developed and validated by Riffo et al. (2013). The test evaluates three main aspects of reading comprehension, namely: (1) the level of processing required by the task, considered "textual" comprehension; (2) the context, also called "pragmatic" understanding; and (3) readers and their perspectives on the text and its context, called "critical" understanding. The test incorporates 4 texts selected according to the school level, namely, a dramatic text (play), a descriptive text (technical text) and two informative texts (film review and news). Each one must be read to answer 30 multiple-choice questions that assess the different levels of processing involved and dimensions of understanding. Sample items include: "Which of the following motive(s) can explain the behavior of Brutus" (critical

understanding of the dramatic text); "Shakespeare and Quevedo are mentioned in the text for the purpose of …" (pragmatic understanding of the dramatic text); and "Towards the end of the text, the author references the popularity of Fausto because …" (textual comprehension of the technical text). Responses are subsequently coded as correct (1) or incorrect (0), and these values are summed to obtain a global score. Next, these raw scores are transformed to percentiles to facilitate interpretation. Internal consistency reliability coefficient, KR-20, for the sample on this measure was 0.793.

For the purposes of data analyses, we conducted a median split procedure to determine proficient readers (reading comprehension score ≥ 43) and poor readers (reading comprehension score ≤ 42; *Min*. = 2; *Max*. = 76; *Med*. = 42). Data were also re-analyzed using extreme scores on either end (i.e., the bottom and top 25% of the distribution) to mitigate the potential bias that may be created by the median split procedure (Tabachnick & Fidell, 2013). It is important to note that the results were nearly identical using both approaches, and hence, we opted to retain the median split approach.

Reading and writing complexity

Reading comprehension and written production were measured using *Trunajod* originally developed and validated by Palma et al. (2019). The innovation of *Trunajod* is that it employs a sophisticated artificial intelligence to automatically analyze any text, and it produces various indices of textual complexity (e.g., sentence length, clause length, lexical diversity, etc.). These different micro linguistic elements can be captured for both reading comprehension and written production by evaluating the complexity of the texts and the written production. Thus, participants receive a detailed score sheet that details their performance on each individual micro linguistic index for both reading comprehension and written production (see Soto et al., 2019; Soto et al., 2020, for the scoring manual and a more in-depth description of each of the indices captured by *Trunajod*).

Writing production

After responding to the reading comprehension test, each participant was asked to write two summaries of two texts belonging to it. The first corresponds to a film review and the second corresponds to a news item. Participants at the time of writing the abstracts could access the texts to reread and obtain information.

The summaries were evaluated by four experts, using an evaluation rubric that is commonly used by the Chilean educational system to rate the written production of students in the fifth grade (i.e., it is a standardized, norm-referenced assessment). This instrument measures mastery of the subject, coherence and progression of ideas, use of varied vocabulary and sentence structure, and mastery of grammar rules.

The rubric was used to comprehensively evaluate the summaries created by the students and had a scale that determined the achievement standards for each indicator. The overall score for each item is the simple sum of the results of each of the indicators. The evaluators were informed about its characteristics, and they were all trained by first reviewing 20 abstracts. With these results, a Cohen's kappa (κ) was conducted for each summary assessment among all possible pairs of rater responses. Once raters had achieved interrater reliability of at least $\kappa = 0.90$, they were asked to rate the missing abstracts. Finally, the

evaluation scores were linearly transformed into percentiles (0-100) to facilitate interpretation and to compare the results with other measures more easily.

Metacognitive monitoring accuracy

We also asked students to provide confidence in performance judgments *before* and *after* the reading and writing tasks on a continuous 0-100 scale, with 0 indicating no confidence in their performance and 100 indicating complete confidence in their performance, as a metric of prospective (before the tasks) and retrospective (after the task) global perceptions of performance (POP), as is common in the metacognitive monitoring literature (e.g., Bol & Hacker 2001; Bol et al., 2005; Gutierrez & Schraw, 2015). We then compared these confidence judgments to their actual reading comprehension and writing performance to obtain a global *absolute* calibration accuracy score (i.e., we calculated the *absolute* difference between confidence in performance judgments and actual performance to avoid potential calculation errors of other composite metacognitive monitoring judgments such as the gamma coefficient [see Schraw et al., 2014 for an explanation of composite monitoring indices]).

Thus, four separate metacognitive monitoring accuracy scores (two global prospective and two global retrospective *absolute* accuracy scores, one each for the reading and writing tasks) served as indices of metacognitive monitoring. Global *absolute* monitoring scores were interpreted such that the higher the score on either task, the greater the error in judgment, and thus, the lower the score the greater the accuracy, with a score of "0" representing perfect accuracy because the signed difference between confidence in performance and actual performance equals "0". While this method produces two possible indices of metacognitive monitoring, accuracy and bias (metacognitive error that can be expressed as overconfidence and underconfidence in performance), we opted to include only accuracy for the present study. This method is typical in the metacognitive monitoring literature and is known as the *residual score approach* (see Schraw 2009, for a discussion on measurement in metacognitive monitoring research).

Procedure

All ethical guidelines for conducting research involving human participants were followed and ethics committee approval was obtained prior to data collection. The *Lectum* reading comprehension test and text production were administered in two sessions, one for each instrument, and the approximate time of each session was 90 minutes, both instruments were administered in groups and online, each participant was located in front of a computer, inserting an empty space between each to facilitate concentration in the computer room of each establishment.

First, the *Lectum* reading comprehension test was applied. Participants were instructed to read the texts to understand the information as best as possible, as after reading each text, they would have to answer comprehension questions. They could neither access other web pages once the test started, nor could they use their cell phones. Participants answered 30 multiple-choice questions from four different texts and, before and after taking the test, they were also asked about their confidence in performance judgments regarding the result using

Google Forms. The question was the following: "From 1 to 100, how well do you think you will do/did on the comprehension test?"

In the second session, a text production guide was administered through Google Forms. Two texts from the reading comprehension test (that the participants had read in the previous session) were selected. Participants were instructed that they should read the texts to understand the information as best as possible, as after reading each text they would have to write a summary of each one, and that their productions should have more than 150 words. They could neither access other web pages once the text production guide had started, nor could they use their cell phones. The texts corresponded to a film review and a newspaper article. Before and after writing each summary, students were asked about their confidence in performance judgments regarding the outcome of their summaries. The question was the following: "From 1 to 100, how good will you/did you perform on your summaries?"

Subsequently, each instrument was evaluated according to its respective rubric. In the case of *Lectum*, with the performance scale provided by the test, in the case of text production, with a rubric provided by the Ministry of Education. The evaluations of each of the written productions were rigorously scored by four teachers belonging to the participating schools. The teachers evaluated the summaries of students from other establishments to avoid biases that could exist with the interpretation of the instruments.

Finally, the summaries of each participant were analyzed by *Trunajod*, where the linguistic indices and the differences within each of the texts produced by the participants were obtained. Once all the data were collected, they were transferred to an EXCEL file for further statistical analysis.

Data Analysis

Data were evaluated for univariate normality using skewness and kurtosis values and histograms with normal curve overlay and multivariate normality via a normalized estimate of the linear combination of dependent variables (Tabachnick & Fidell, 2013). All variables approximated univariate and multivariate normality across both proficient and poor readers. No cases were classified as outliers through box-and-whisker plots by group (univariate) and via review of Cook's *D* and Mahalanobis Distance (multivariate), and thus, all 120 cases were retained for analysis. There were no missing data, as all participants completed all data points. Other assumptions such as homoscedasticity, homogeneity of variance (univariate) and homogeneity of variance-covariance matrices (multivariate), and lack of multicollinearity were also met. Therefore, we proceeded with the planned analyses without making any adjustments to the data. The Bonferroni adjustment to statistical significance was employed to control familywise Type I error rate inflation. All data were analyzed via IBM's Statistical Package for the Social Sciences (SPSS) 23.

The first research question was answered by conducting a doubly multivariate repeated measures analysis of variance (RM MANOVA). Reading comprehension proficiency (proficient, poor) served as the fixed between-subjects factor and type of metacognitive monitoring judgment (POP; prospective, retrospective) for reading and writing served as the within-subjects factors. The second research question was answered by conducting a series of ordinary least squares (OLS) hierarchical regressions, one each for proficient and poor readers separately. In each of the two hierarchical regressions, the *Trunajod* indices of linguistics served as predictors and retrospective metacognitive monitoring accuracy in the

Table 1 Descriptive Statistics for the Sample and for Proficient and Poor Readers

Variable	Sample		Proficient ($n=55$)		Poor ($n=65$)	
	M	SD	M	SD	M	SD
Reading Comprehension	48.05	19.84	58.24	12.17	23.63	8.21
Reading Monitoring 1	44.54	28.92	27.34	25.67	56.83	24.63
Reading Monitoring 2	30.25	30.65	11.40	26.96	43.71	25.74
Writing Monitoring 1	23.33	30.93	16.90	29.37	27.91	31.40
Writing Monitoring 2	28.16	30.83	21.54	28.68	32.88	31.63
CL-PT Writing Total	35.86	17.61	41.90	18.33	31.54	16.00

$N=120$

Key. 1 represents prospective (before the test) and 2 represents retrospective (after the test) monitoring accuracy judgments

Table 2 Zero-Order Correlation Matrix of Measures for Proficient and Poor Readers

Variable	1	2	3	4	5	6
1. Reading Comprehension	-	−0.43**	−0.52**	−0.11	−0.13	0.20
2. Reading Monitoring 1	−0.70**	-	0.68**	0.10	0.16	−0.10
3. Reading Monitoring 2	−0.59**	0.60**	-	0.19	0.19	−0.11
4. Writing Monitoring 1	−0.10	0.20	0.04	-	0.81**	−0.53**
5. Writing Monitoring 2	−0.18	0.19	0.39**	0.60**	-	−0.52**
6. CL-PT Writing Total	0.31*	−0.28*	−0.05	−0.69**	−0.63**	-

Note. Correlations above the diagonal are for poor readers ($n=65$) and those below the diagonal are for proficient readers ($n=55$)

$N=120$ * $p<.05$ ** $p<.01$ (one-tailed test of significance)

writing task served as the criterion while controlling for the effect of prospective monitoring accuracy. Effect sizes for the MANOVA were reported as partial η^2 (η^2_p) and those for the regression analyses were reported as the R^2. Cohen (1988) provided the following interpretive guidelines for η^2_p: 0.010-0.059 as small; 0.060-0.139 as moderate; and ≥0.140 as strong. For R^2, these values were: 0.010-0.249 as small; 0.250-0.499 as moderate; and ≥0.500 as strong.

Results

Descriptive statistics by group are displayed in Tables 1 and 2 presents zero-order bivariate correlations by group.

Descriptive statistics in Table 1 show that, generally, proficient readers not only exhibited greater performance in reading comprehension and writing than poor readers, but that proficient readers were also more accurate in their metacognitive monitoring judgments than poor readers. Zero-order bivariate correlations in Table 2 indicate that all relations were in the theoretically expected direction. The negative correlations with the metacognitive monitoring accuracy variables in reading and writing are expected, given that we employed the residual score approach in calculating monitoring accuracy (i.e., higher scores revealed greater errors in judgment). It is also interesting to note that, in general, the correlation coefficients were higher for proficient readers than for poor readers.

RQ1: Differences between proficient and poor readers in reading and writing monitoring accuracy

Results of the doubly multivariate RM MANOVA revealed that none of the group x type of monitoring judgment interactions reached statistical significance, all p-values ≥ 0.485. Nevertheless, the type of judgment main effect reached statistical significance for reading, $F(1,118)=51.99$, $p<.001$, $\eta^2_p=0.306$, and writing, $F(1,118)=6.38$, $p=.012$, $\eta^2_p=0.054$. The reading proficiency main effect also reached significance for reading, $F(1,118)=51.52$, $p<.001$, $\eta^2_p=0.304$, and writing, $F(1,118)=8.53$, $p=.001$, $\eta^2_p=0.067$.

Regarding the type of judgment main effect, both proficient and poor readers exhibited better monitoring accuracy in retrospective judgments (after the task) than in prospective judgments (before task) for both the reading and writing task. With respect to the reading proficiency main effect, proficient readers demonstrated significantly more accurate monitoring compared to poor readers in both reading and writing tasks.

RQ2: Differences in predictive patterns of linguistic indices on writing monitoring accuracy between proficient and poor readers

Poor readers

Results of the hierarchical linear regression analysis for poor readers revealed that, after controlling for the effect of prospective monitoring accuracy in writing ($b=0.90$, 95% confidence interval=0.64, 1.16, $\beta=0.87$), the model with *function_ttr* (functional word diversity) ($b=86.71$, 95% confidence interval=49.41, 124.01, $\beta=0.41$) and *density_frase nominal* (noun phrase density) (b = -30.28, 95% confidence interval = -56.43, -4.13, $\beta = -0.20$) significantly predicted writing retrospective monitoring accuracy, $\Delta F(3,62)=6.09$, $\Delta p=0.02$, $\Delta R^2=0.153$.

Poor readers have the peculiarity that when they write they support monitoring in their composition process based on certain superficial indices of what is written, and they are not necessarily variables that help to configure a richer mental representation. In fact, from a lexical point of view, functional words are not content words but function words; that is, they act by linking the words, but do not add relevant information about what is written. Additionally, they use the density of the noun phrases as a key, which could act as a sign of the complexity of what they are writing.

Proficient readers

For proficient readers, results of the hierarchical linear regression analysis showed that, after controlling for the effect of prospective monitoring accuracy in writing ($b=1.06$, 95% confidence interval=0.55, 1.58, $\beta=0.86$), the model with *disimilaridad PoS por oración* (similarity/dissimilarity between sentence) ($b=182.14$, 95% confidence interval=156.13, 208.14, $\beta=0.57$), *diversidad léxica MTLD* (lexical diversity) ($b=-0.49$, 95% confidence interval=-0.67, -0.32, $\beta=-0.27$), and *imageability* (use of words that readily lend themselves to be visualized) ($b=8.83$, 95% confidence interval=3.27, 14.38, $\beta=0.15$) signifi-

cantly predicted writing retrospective monitoring accuracy, $\Delta F(4,51)=12.22$, $\Delta p<0.01$, $\Delta R^2=0.258$.

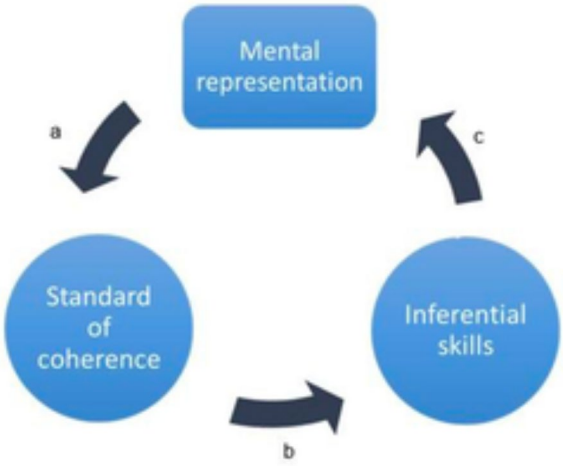
Proficient readers use a more relevant mix of cues, which help in their process of monitoring their writing. Unlike poor readers, they use as cues the diversity of the content words of writing, the similarity between the sentences, and the use of words that readily lend themselves to be visualized. The latter is relevant because it helps to facilitate the precision with which the mental representation of what is written is generated.

Discussion

The aim of this study was to examine the influence of reading comprehension proficiency (proficient, poor) on elementary school students' metacognitive monitoring accuracy in reading and writing tasks. Results indicate that participants improve the accuracy of their monitoring after the task is done (retrospective monitoring) in both reading and writing, especially proficient readers. This finding is supported by extant research on metacognitive monitoring that converges on the conclusion that learners tend to be more accurate in their retrospective performance judgments (after the task) than their prospective judgments (before the task; e.g., Gutierrez et al., 2016; Hadwin & Webster, 2013). However, the story becomes more interesting regarding the relationship between retrospective monitoring accuracy of reading and writing for proficient readers, which is an innovative and interesting finding. Proficient readers employ different types of cues which are linked to their monitoring accuracy in writing. Among them are lexical diversity, type of words, and syntactic variables. This supports findings from a previous study, in which a greater variety of linguistic cues were associated with the writing skill of proficient readers (Soto et al., 2020). Additionally, this study and the one by Gutierrez de Blume et al. (2021) show a consistent cue, use of words that readily lend themselves to be visualized, employed by proficient readers, serves as a key element in their written composition skill. Apparently, proficient readers use words of this type to improve their textual production, which allows them to enrich their elaboration process. Conversely, for poor readers, the cues in their writing skill have to do with elements that are more focused on a surface level of the text, both at the word and sentence level.

The latter is closely related to the most important finding of this study. As readers improve their reading performance, so, too, does writing, not only in terms of overall performance, but in terms of metacognitive monitoring accuracy. As different investigations have shown (Baker, 1984, 1996; Oakhill et al., 2005; Yang, 2002), as reading proficiency rises, monitoring accuracy increases with it, which shows that the type of cues that readers rely on to estimate their judgments of performance are coinciding (Soto et al., 2019). A potential explanatory path to these findings was provided by Gutierrez de Blume et al. (2021). This study found that as readers' mental representation is enriched, the relations between reading comprehension and writing production become stronger. Presumably, this is because readers begin to rely on more consistent (and relevant) cues regarding the level of their performance. According to what Gutierrez de Blume et al. (2021) argued, as inferential processes improve, the mental representation of what individuals write or read improves. As mental representation becomes richer, the level of demand in establishing the standard of coherence required to accept what is represented (both in reading and writing) increases.

Fig. 1 *Cycle of Learning Regulation*



When the standard is not achieved, inferential processes are activated, which recalibrate the cycle of mental representation on the object to be represented. Evidently, this cycle has an important effect on learning, which is determined by the joint action of cognitive and metacognitive processes that are related from a learning regulation cycle, which includes three components: (1) standard of coherence; (2) mental representation; and (3) inferential skills (Gutierrez de Blume et al., 2021).

Implications and recommendations for Research, Theory, and practice

Our work contributes to a better understanding of the problem of calibration accuracy in reading comprehension and written production in school children. Everything indicates, on the one hand, that when competent readers and writers present a better mental representation of the text, they use cues that are linked to a more precise follow-up in reading and writing tasks and that they use linguistic indices that support the monitoring of their writing. That is, readers with more developed skills produce better texts and have better management of their metacognitive skills. Due to the above, it is necessary to strengthen the mental representation and with this raise reading skills to improve writing; this could be a "crossroads" because this happens mainly through reading. Pedagogy plays a key role here and, within it, teaching action. Teachers should be motivated to develop practices aimed at strengthening the development of skills that allow activating the parameter of coherence standards such as inferential skills. This involves a sequence of alternative modeling and training exercises. Surely modeling should focus on capturing coherence gaps, while training should allow those gaps to be narrowed. We believe that this intervention would improve the cycle of the three dimensions (see Fig. 1), and this improvement would lead to better learning outcomes.

A high standard of coherence increases the formation of inferences, and this increase permits the development of a more coherent mental representation, forming a continuous cycle of learning. The difference between the current mental representation one has and its standard of coherence regulates the activation of new inferential processes required to

enhance the mental representation. When the processes related to the standard of coherence and inference-generation are improved, the result is a better mental representation. Continued investigation regarding this model and pedagogical techniques can enrich interventions designed to enhance different points during these processes. First, researchers could evaluate the relation between standard of coherence and mental representation. This includes the cognitive-metacognitive operations where individuals are likely to be aware of their representation and other possible representations, enriching their standard of acceptance. Second, the gaps in coherence are decreased when a metacognitive-cognitive process starts based on the evaluation of the mental representation one has by its standard of coherence to prompt inferential production. Finally, we propose that an intervention at some point in the process could alter the learning cycle, such as the inference operations, which could enrich the mental representation.

However, what would happen if readers generated a high standard of coherence, but lack the inferential skills to improve their representation? Plausibly, students could generate frustration because they lack the means to satisfy that standard. On the other hand, if readers generate many inferences, but they are not able to adequately monitor and evaluate their mental representation, they are incapable of deciding if they are generating a precise level of representation, disabling the adjustment process. Hence, this new conceptualization assumes cycles of learning based on the components mentioned, where the focus is the mental representation. Working on that cycle and the sub-processes presented could inform the development of different pedagogical interventions, generating focused, finer-grained activities, materials, and sequences of tasks necessary to improve the subjacent processing of learning.

Our work places the coherence standard, mental representation, and inferential skills as the three primary components that are involved in the learning process, understanding our perspective as a wider framework where cognitive and metacognitive processes converge, which is especially useful to generate pedagogical interventions. Thus, according to the results of this study, there are two different ways the cycle of learning regulation for proficient and poor readers/writers manifest. The proficient learner, both in reading and writing, implements inferential operations that help to generate a high level of mental representation. Next, that high mental representation is contrasted with the coherence standard to implement inferential operations again. We assume that this iterative cycle ensures a high quality of the representation.

Avenues for Future Research

One fruitful future avenue could be to generate research processes that address the monitoring of reading and writing around the same *objects of knowledge* to better understand the relationship between both dimensions of literacy. This will allow us to isolate prior knowledge or other variables that affect or blur the findings around the relationship between reading, writing, and mental representation. Another viable research program is to combine all or multiple aspects of metacognition such as conditional knowledge and planning, information management, and evaluation to observe their relation to reading and writing tasks between proficient and poor readers. This may even permit the experimental manipulation

of metacognitive skills training and its effects on the reading and writing performance of proficient and poor readers.

Methodological reflections and Limitations

No research study is without limitations. The present study employed a non-experimental research design, which limits the inferences and conclusions that can be reasonably drawn from the data. Likewise, one cannot infer any causal effects from the findings of the present study. Finally, the sample size of the present investigation may be considered relatively small by some, and thus, results should be interpreted with caution. It is also important to note that, because of the nature of *Lectum* and the written production tasks, some participants may have felt anxious regarding performance, which may have influenced the findings.

Despite these limitations, however, the present study employed objective measures instead of subjective ones like self-report surveys. The use of objective rather than subjective measures enhances the claims and conclusions one can reasonable extract from the findings. Further, the effect sizes of the present study can be considered robust, including moderate to large effects, thereby minimizing the likelihood of unstable or spurious effects. Finally, this study explored the relation between reading comprehension, written production, and metacognitive monitoring accuracy, which, to our understanding, has not yet been found in extant research. Therefore, the present study provides a meritorious contribution to the research literature on the role of metacognitive monitoring on reading comprehension and written production as a function of reading proficiency.

Conclusions

Key findings of the present study indicated that skilled readers exhibited more accurate metacognitive monitoring than less skilled readers. Additionally, regression models showed that skilled readers engaged more frequently in analyzing deeper aspects of texts (i.e., reading comprehension), which improved comprehension, and were able to use more strategic metacognitive monitoring cues to enhance their writing production. Skills involved included inferencing, paraphrasing, bridging, and focusing on main ideas. Poor readers, on the other hand, relied on superficial aspects of texts, including memorization. Thus, the take home message that the present study tentatively makes is that the metacognitive cues that proficient readers employ during reading comprehension and writing production appear to be more sophisticated whereas poor readers tend to use more superficial, text-base cues. This aligns closely with key guidelines from Kintsch's (1988) C-I model and Otero's (2002) *regulation model*, which, as the present study suggests, help explain how reading comprehension improves metacognitive monitoring during reading and writing production, which leads to appreciable gains in learning. Thus, innovative educational interventions should be tailored to the improvement of metacognitive monitoring, reading comprehension, and writing production of individuals with underdeveloped reading skills using some of the lessons gleaned from this study.

Funding This research was supported by a FONDEF ID grant, number 20I10290, titled "Literador: Un tutor inteligente que potencia las competencias en lectura y escritura" [Literador: An intelligent tutor that improves reading and writing skills], from ANID, Chile.

Data Availability Data associated with this research are available to anyone upon reasonable request.

Code Availability Not applicable.

Declarations

Conflicts of interest/Competing interests: None of the authors have any conflicts of interest to report.

References

Baker, L. (1979). Comprehension monitoring: Identifying and coping with text confusions. *Journal of Reading Behavior*, *11*, 363–374. https://doi.org/10.1080%2F10862967909547342

Baker, L. (1984). Spontaneous versus instructed use of multiple standards for evaluation comprehension: Effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology*, *38*, 289–311. https://doi.org/10.1016/0022-0965(84)90127-9

Baker, L. (1996). Social influence on metacognitive development in reading. In C. Cornoldi & J Oakhill, (Ed.), *Reading comprehension difficulties* (pp.331–351). Lawrence Erlbaum Associates

Beal, C. R., Garrod, A. C., & Bonitatibus, G. J. (1990). Fostering children's revision skills through training in comprehension monitoring. *Journal of Educational Psychology*, *82*(2), 275–280. https://doi.org/10.1037/0022-0663.82.2.275

Beauvais, C., Olive, T., & Passerault, J. M. (2011). Why are some texts good and others not? Relationship between text quality and management of the writing processes. *Journal of Educational Psychology*, *103*(2), 415–428. https://doi.org/10.1037/a0022545

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Lawrence Erlbaum Associates, Inc

Berninger, V. (2000). Development of language by hand and its connections to language by ear, mouth, and eye. *Topics of Language Disorders*, *20*, 65–84. https://doi.org/10.1097/00011363-200020040-00007

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education*, *69*(2), 133–151. https://doi.org/10.1080/00220970109600653

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

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68–86. https://doi.org/10.1016/j.jml.2015.12.009

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*(4), 671. https://doi.org/10.1037/0022-0663.96.4.671

Chen, M. H., Gualberto, P. J., Tameta, C. L., & Salle, D. L. (2009). The development of metacognitive reading awareness inventory. *TESOL Journal*, *1*(1), 43–57

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The tool for the automatic analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavior Research Methods*, *51*(1), 14–27. https://doi.org/10.3758/s13428-018-1142-4

Dunlosky, J., Griffin, T., Thiede, K., & Wiley, J. (2005). Understanding the delayed keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning Memory & Cognition*, *31*, 1267–1280. https://doi.org/10.1037/0278-7393.31.6.1267

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *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., & Hacker, D. J. (2002). Metacomprehension of science text: Investigating the levels-of-disruption hypothesis. In J. Otero, J. A. León, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 255–279). Lawrence Erlbaum Associates Publishers

Ehrlich, M. F. (1996). Metacognitive monitoring in the processing of anaphoric devices in skilled and less-skilled comprehenders. In C. Cornoldi, & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processesand remediation* (pp. 221–249). Erlbaum

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*, *11*, 29–63. https://doi.org/10.1023/A:1007996502372

Ferrari, M., Bouffard, T., & Rainville, L. (1998). What makes a good writer? Differences in good and poor writers' self-regulation of writing. *Instructional Science*, *26*(6), 473–488. https://doi.org/10.1023/A:1003202412203

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, *35*(1), 39–50. https://doi.org/10.1207/S15326985EP3501_5

Flavell, J. H. (1978). Metacognitive development. In J. M. Scandura, & C. J. Brainerd (Eds.), *Structural/process theories of complex human behavior* (pp. 34–78). Sijthoff & Noorddhoff

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365–387. https://doi.org/10.2307/356600

Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production-comprehension link. *Journal of memory and language*, *89*, 110–137. https://doi.org/10.1016/j.jml.2015.10.001

Gernsbacher, M. A., & Givón, T. (Eds.). (1995). *Coherence in spontaneous text* (31 vol.). John Benjamins Publishing

Grabe, W., & Kaplan, R. B. (2014). *Theory and practice of writing: An applied linguistic perspective*. Routledge

Graesser, A. C., McNamara, D. S., Louwerse, M. M., et al. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods Instruments & Computers*, *36*, 193–202. https://doi.org/10.3758/BF03195564

Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, *81*(4), 710–744. https://doi.org/10.17763/haer.81.4.t2k0m13756113566

Graham, S., Liu, X., Aitken, A., Ng, C., Bartlett, B., Harris, K. R., & Holzapfel, J. (2018). Effectiveness of Literacy Programs Balancing Reading and Writing Instruction: A Meta-Analysis. *Reading Research Quarterly*, *53*(3), 279–304. https://doi.org/10.1002/rrq.194

Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., Barkel, A., Kavanaugh, C., & Talukdar, J. (2018). Reading for Writing: A Meta-Analysis of the Impact of Reading Interventions on Writing. *Review of Educational Research*, *88*(2), 243–284. https://doi.org/10.3102/0034654317746927

Gutierrez, A. P., & Price, A. F. (2017). Calibration between undergraduate students' prediction of and actual performance: The role of gender and performance attributions. *The Journal of Experimental Education*, *85*, 486–500. https://doi.org/10.1080/00220973.2016.1180278

Gutierrez, & Schraw, G. (2015). Effects of strategy training and incentives on students' performance, confidence, and calibration. *Journal of Experimental Education*, *83*(3), 386–404. https://doi.org/10.1080/00220973.2014.907230

Gutierrez de Blume, A. P. (2017). The effects of strategy training and an extrinsic incentive on fourth- and fifth-grade students' performance, confidence, and calibration accuracy. *Cogent Education*, *4*(1), 1–17. https://doi.org/10.1080/2331186X.2017.1314652

Gutierrez de Blume, A. P. (2020). Efecto de la instrucción de estrategias en la precisión del monitoreo metacognitivo de los alumnos universitarios estadounidenses. *Tesis Psicológica, 15*(2), 1–29. https://doi.org/10.37511/10.37511/tesis.v15n2a9

Gutierrez, A. P., Schraw, G., Kuch, F., & Richmond, A. S. (2016). A two-process model of metacognitive monitoring: Evidence for general accuracy and error factors. *Learning and Instruction*, *44*, 1–10. https://doi.org/10.1016/j.learninstruc.2016.02.006

Gutierrez de Blume, A. P., Soto, C., Ramírez Carmona, C., Rodriguez, F., & Castillo, P., P (2021). Reading competence and its impact on writing: an approach towards mental representation in literacy tasks. *Journal of Research in Reading*, *44*(3), 617–635. https://doi.org/10.1111/1467-9817.12359

Guzzardo Tamargo, R. E., Valdés Kroff, J. R., & Dussias, P. E. (2016). Examining the relationship between comprehension and production processes in code-switched language. *Journal of Memory and Language*, *89*, 138–161. https://doi.org/10.1016/j.jml.2015.12.002

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky, & R. Bjork (Eds.), *Handbook of Memory and Metacognition* (pp. 429–455). Erlbaum

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (1998). *). Metacognition in educational theory and practice*. Routledge

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

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, *53*(4), 449. https://doi.org/10.1037/0003-066X.53.4.449

Harris, K. R., & Graham, S. (1992). Self-regulated strategy development: A part of the writing process. In M. Pressley, K. R. Harris, & J. T. Guthrie (Eds.), *Promoting academic competence and literacy in school* (pp. 277–309). Academic Press

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing: An Interdisciplinary Approach.* (pp. 3–30). Lawrence Erlbaum Associates. https://doi.org/BF 456 W8 C676 1980

Honeycutt, R. L. (2002). *Good reader/poor writer: An investigation of the strategies, understanding, and meaning that good readers who are poor writers ascribe to writing narrative text on-demand*. http://www.lib.ncsu.edu/resolver/1840.16/5094

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Hsiao, Y., & MacDonald, M. C. (2016). Production predicts comprehension: Animacy effects in Mandarin relative clause processing. *Journal of Memory and Language*, *89*, 87–109. https://doi.org/10.1016/j.jml.2015.11.006

Jacovina, M., & McNamara, D. S. (2017). Intelligent tutoring systems for literacy: Existing technologies and continuing challenges. In R. Atkinson (Ed.), *Intelligent tutoring systems: Structure, applications, and challenges*. Nova Science Publishers Inc

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction*, *7*(3), 161–195. https://doi.org/10.1207/s1532690xci0703_1

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

Kittredge, A. K., & Dell, G. S. (2016). Learning to speak by listening: Transfer of phonotactics from perception to production. *Journal of Memory and Language*, *89*, 8–22. https://doi.org/10.1016/j.jml.2015.08.001

Lee, C. (2004). Seeing is understanding: Improving coherence in students' writing. *The Internet TESL Journal*, *10*(7), 247–265. https://doi.org/10.2307/3586543

Lika, M. (2017). Teaching reading comprehension strategies. *Academic Journal of Business Administration Law and Social Sciences*, *3*(1), 196–204

Logan, S., Medford, E., & Hughes, N. (2011). The importance of intrinsic motivation for high and low ability readers' reading comprehension performance. *Learning and Individual Differences*, *21*(1), 124–128. https://doi.org/10.1016/j.lindif.2010.09.011

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning Memory and Cognition*, *10*(4), 663–679. https://doi.org/10.1037/0278-7393.10.4.663

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*(4), 723–731. https://doi.org/10.1037/0022-0663.97.4.723

McNamara, D. S. (Ed.). (2007). *Reading comprehension strategies: Theories, interventions, and technologies*. Psychology Press

Meyer, A. S., Huettig, F., & Levelt, W. J. M. (2016). Same, different, or closely related: What is the relationship between language production and comprehension? *Journal of Memory and Language*, *89*, 1–7. https://doi.org/10.1016/j.jml.2016.03.002

National Assessment of Educational Progress (NAEP). (2003). *Trends in academic progress*. National Center for Education Statistics, U.S. Department of Education. https://nces.ed.gov/nationsreportcard/

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, *18*, 657–686. https://doi.org/10.1007/s11145-005-3355-z

OECD (2017). PISA 2015 Results

Otero, J. (2002). Noticing and fixing difficulties while understanding science texts. In J. Otero, J. A. León, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 281–308). Lawrence Erlbaum Associates

Palma, D., & Soto, C. (2022). Open-Source Collaboration to Assess the Text Complexity, Helping to Read and Write in Spanish-Speaking Schools. *Human Systems Engineering and Design (IHSED 2021)*.https://doi.org/10.54941/ahfe1001171

Palma, D., Soto, C., Veliz, M., Karelovic, B., & Riffo, B. (2021). TRUNAJOD: A text complexity library to enhance natural language processing. *Journal of Open Source Software*, *6*(60), 3153. https://doi.org/10.21105/joss.03153

Palma, D., Soto, C., Veliz, M., Riffo, B., & Gutiérrez, A. (2019). A Data-Driven Methodology to Assess Text Complexity Based on Syntactic and Semantic Measurements. *International Conference on Human Interaction and Emerging Technologies*, 509–515. Springer, Cham

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling, & C. Hume (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford, England: Blackwell. https://doi.org/10.1002/9780470757642.ch13

Pintrich, P. R., Wolters, C., & Baxter, G. (2000). Assessing metacognition and selfregulated learning. In G. Schraw, & J. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Lincoln, NE: Buros Institute of Mental Measurement

Riffo, B., Véliz, M., Castro, G., Reyes, F., Figueroa, B., Salazar, O., & Herrera, M. O. (2013). *LECTUM. Prueba de comprensión lectora*. Conicyt. Proyecto Fondef D08i1179

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Schraw, G., Kuch, F., Gutierrez, A. P., & Richmond, A. S. (2014). Exploring a Three-Level Model of Calibration Accuracy. *Journal of Educational Psychology*, *106*(4), 1192–1202. https://doi.org/10.1037/a0036653

Smith, K. S., Aziz, W., & Specia, L. (2016). Cohere: A toolkit for local coherence. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 4111–4114

Soto, C., Gutiérrez de Blume, A. P., Carrasco Bernal, M. A., & Contreras Castro, M. A. (2020). The role of meta-cognitive cues on the comprehension of proficient and poor readers. *Journal of Research in Reading*, 1–18. https://doi.org/10.1111/1467-9817.12303

Soto, C., Gutiérrez de Blume, A. P. G., Castro, M. A. C., & Bernal, M. C. (2019). Development and Validation of a Tool to Detect and Repair Text Inconsistencies. *Electronic Journal of Research in Educational Psychology*, *17*(48), 437–464. https://doi.org/10.25115/ejrep.v17i48.2352

Soto, C., Gutiérrez de Blume, A. P., Jacovina, M., McNamara, D., Benson, N., Riffo, B., & Kruk, R. (2019). Reading comprehension and metacognition: The importance of inferential skills. *Cogent Education*, *6*(1), https://doi.org/10.1080/2331186X.2019.1565067

Soto, C., Gutierrez de Blume, A. P., Rodríguez, M. F., Asún, R., Figueroa, M., & Serrano, M. (2019). Impact of bridging strategy and feeling of knowing judgments on reading comprehension using COMPRENDE: An educational technology. *TechTrends*, *63*(5), 570–582. https://doi.org/10.1007/s11528-019-00383-5

Soto, C., Rodríguez, M. F., & de Blume, A. P. G. (2018). Exploring the meta-comprehension abilities of students with intellectual disabilities. *International Journal of Special Education*, *33*(2), 233–247

Spiegel, D. L., & Fitzgerald, J. (1990). Textual cohesion and coherence in children's writing revisited. Research in the Teaching of English,48–66. https://www.jstor.org/stable/40171445

Stone, N. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, *12*, 437–475. https://doi.org/10.1023/A:1009084430926

Struthers, L., Lapadat, J. C., & MacMillan, P. D. (2013). Assessing cohesion in children's writing: Development of a checklist. *Assessing Writing*, *18*(3), 187–201. https://doi.org/10.1016/j.asw.2013.05.001

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*, *47*(4), 331–362. https://doi.org/10.1080/01638530902959927

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (pp. 85–106). Routledge/Taylor & Francis Group

UNESCO (2015). *Informe nacional de resultados TERCE*. Extraído de http://archivos.agenciaeducacion.cl/TERCE_informefinal.pdf

Weaver, C. (1990). *Understanding whole language: From principles to practice*. Heinemann

Weaver, C. A., & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, *23*(1), 12–22. https://doi.org/10.3758/BF03210553

Westby, C. (2004). A language perspective on executive on functioning, metacognition, and self-regulation. In A. Stone, E. R. Silliman, B. J. Ehren, & K. Apel (Eds.), *Handbook of language and literacy* (pp. 398–427). Guilford Publication

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology*, *132*(4), 408–428. https://doi.org/10.3200/GENP.132.4.408-428

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication*, *32*, 189–204. https://doi.org/10.2307/356693

Wong, B. Y. (1999). Metacognition in writing. In R. Gallimore, L. P. Bernheimer, D. L. McMillan, D. L. Speece, & S. R. Vaughn (Eds.), *Developmental perspectives on children with high-incidence disabilities* (pp. 183–198). Routledge

Yang, Y. F. (2002). Reassessing readers' comprehension monitoring.Reading in a Foreign Language, *14*(1). http://nflrc.hawaii.edu/rfl/April2002/yang/yang.html

Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education*, *23*(1), 31–48. https://doi.org/10.1016/j.linged.2011.09.004

Zamuner, T. S., Morin-Lessard, E., Strahm, S., & Page, M. P. A. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language*, *89*, 55–67. https://doi.org/10.1016/j.jml.2015.10.003

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Taylor & Francis Group. https://doi.org/10.4324/9780203771587