Mesure et évaluation en éducation

Validating a Scale Measuring Metacognitive Knowledge About Digital Reading

Dominique Lafontaine, Nina Jude and Johanna Leck

Volume 45, Number spécial, 2022

Translation Issue

Acceptance: January 11, 2022
Final version: June 28, 2022
Acceptance: September 21, 2022

URI: https://id.erudit.org/iderudit/1108759ar
DOI: https://doi.org/10.7202/1108759ar

See table of contents

Publisher(s)

ADMEE-Canada

ISSN

0823-3993 (print)
2368-2000 (digital)

Explore this journal

Cite this article

Lafontaine, D., Jude, N. & Leck, J. (2022). Validating a Scale Measuring Metacognitive Knowledge About Digital Reading. *Mesure et évaluation en éducation, 45*(spécial), 63–94. https://doi.org/10.7202/1108759ar

Article abstract

This study aims to find evidence that validates an unprecedented scale assessing metacognitive knowledge about digital reading, developed for PISA 2018. More precisely, the study investigates whether the scale measures metacognitive knowledge about reading as opposed to ICT literacy. Data from the 37 OECD countries were used. The results show strong correlation between this metacognitive knowledge scale, reading performance, and two scales which have already been validated for measuring metacognition about reading. The scale is also significantly correlated with students' interest, self-concept, and self-efficacy in reading. On the other hand, no correlation was shown between their use of ICT, and their interest and perceived self-efficacy in their use. The results provide evidence that the scale assesses metacognitive knowledge about digital reading rather than digital literacy. The results also demonstrate a strong correlation between the scale and reading proficiency.

Validating a Scale Measuring Metacognitive Knowledge About Digital Reading[1]

Dominique Lafontaine
ORCID 0000-0003-1497-4634
University of Liège

Nina Jude
ORCID 0000-0001-6755-0435
Heidelberg University

Johanna Leck
Stiftung Lesen (Reading Foundation)

Keywords: metacognitive awareness, digital reading, validation process, large-scale assessments

This study aims to find evidence that validates an unprecedented scale assessing metacognitive knowledge about digital reading, developed for PISA 2018. More precisely, the study investigates whether the scale measures metacognitive knowledge about reading as opposed to ICT literacy. Data from the 37 OECD countries were used. The results show strong correlation between this metacognitive knowledge scale, reading performance, and two scales which have already been validated for measuring metacognition about reading. The scale is also significantly correlated with students' interest, self-concept, and self-efficacy in reading. On the other hand, no correlation was shown between their use of ICT, and their interest and perceived self-efficacy in their use. The results provide evidence that the scale assesses metacognitive knowledge about digital reading rather than digital literacy. The results also demonstrate a strong correlation between the scale and reading proficiency.

1 The authors are responsible for the translation process, ensuring proper alignment between the French and English versions. The French version was published in issue 45(1) 2022: https://doi.org/10.7202/1097151ar.

Mots-clés: connaissances métacognitives, lecture numérique, processus de validation, enquêtes à large échelle

Cette étude vise à valider une échelle de mesure des connaissances métacognitives de la lecture numérique inédite, développée pour le PISA 2018 et, plus particulièrement, à établir si cette échelle mesure des connaissances relevant de la lecture plutôt que de la littératie numérique. Les données des 37 pays de l'OCDE ont été utilisées. Les résultats montrent que l'échelle de connaissances métacognitives présente une forte corrélation avec les performances en lecture et avec deux échelles ayant subi un processus de validation de la mesure de la métacognition en lecture. Cette échelle est aussi significativement liée à l'intérêt des élèves, à leur concept de soi et à leur sentiment d'efficacité en lecture, mais n'est pas liée avec leur utilisation des TIC, leur intérêt ou leur sentiment d'efficacité perçue dans ce domaine. Les résultats appuient le fait que l'échelle mesure des connaissances relevant de la lecture plutôt que de la littératie numérique et mettent en évidence une corrélation robuste entre les connaissances métacognitives et les performances en lecture.

Palavras chaves: conhecimento metacognitivo, leitura digital, processo de validação, pesquisas em larga escala

Este estudo tem como objetivo identificar os indícios de validade de u m a escala inédita de medida de conhecimentos metacognitivos da leitura digital, desenvolvida para o PISA 2018 e, mais especificamente, verificar se essa escala mede conhecimentos mais relacionados com a leitura do que com a literacia digital. Foram usados dados de 37 países da OCDE. As correlações mostram que a escala de conhecimentos metacognitivos está fortemente correlacionada com o desempenho em leitura e com duas escalas de medição de metacognição de leitura validadas. Está também significativamente relacionada com o interesse dos alunos, o seu autoconceito e o seu sentimento de eficácia em leitura, mas não está relacionada com a utilização das TIC, o seu interesse ou o seu sentimento de eficácia percebido neste domínio. Os resultados validam o facto de que a escala mede conhecimentos mais relacionados com a leitura do que com a literacia digital e demonstra uma correlação robusta dos conhecimentos metacognitivos com os desempenhos em leitura.

Authors' note: Correspondence regarding to this article may be addressed to dlafontaine@uliege.be

Introduction

Since the 2000s, the Program for International Student Assessment (PISA) has gradually established itself as one of the most influential international assessments of educational policy. The three-yearly PISA program, set up by the Organisation for Economic Co-operation and Development (OECD) in 2000, is a stringent method for assessing the development of knowledge and skills in students from the participating countries[2]. PISA covers three areas: reading, mathematics, and science. Each PISA cycle focuses on one of the three cognitive domains, which is assessed in greater depth. Reading was the focus of PISA 2018, for the third time. PISA has been administered electronically in all OECD countries since 2015.

Most of the media attention on PISA results focuses on country rankings according to student performance. However, PISA also collects contextual information from students and school principals about family environment, learning contexts, school characteristics, and education policies. All this information is used to provide a more detailed analysis of students' achievement (e.g.: by gender or by sociocultural background) and to identify success factors within and between countries. Since 2015, the development of the assessment tests for the three domains and of contextual questionnaires has been based on a conceptual framework (OECD, 2019). This framework specifies the categories and constructs to be investigated in context questionnaires. One of the categories involves the non-cognitive and metacognitive aspects of reading. Thus, concepts such as attitudes, motivation, and strategies related to reading need to be studied in depth.

The present study investigates one of the scales which is designed to measure metacognitive knowledge about digital reading. To our knowledge, this is the first time that a scale specifically targeting metacognitive

2. In 2018, 79 countries took part in PIS: 37 OECD member countries and 42 partner countries or economies.

knowledge about digital reading has been developed in an international context. Given PISA's broad influence, some instruments developed for PISA, which have been validated (e.g. the Interest in Reading scale), have been used in other surveys (notably PIRLS[3]) and several secondary studies (Brozo et al., 2014; Lihong et al., 2021). This scale measuring metacognitive knowledge in a digital reading context, which is rare to date (Burin et al., 2020; Li, 2020), is strongly correlated with reading performance (OECD, 2021). As a result, it is likely to be used in surveys other than PISA, making its validation even more important.

Print and Digital Reading

Recent cognitive models of reading all emphasize the interactive nature of reading: comprehension results from interaction between a reader, a text, and a context (Kintsch & Kintsch, 2005; McNamara & Magliano, 2009; Snow & the RAND Corporation, 2002). Readers construct meaning using their existing knowledge, and use three main types of strategy: (a) locate information; (b) manage and maintain comprehension; and (c) critically assess the validity and relevance of the information. Each of these strategy types can be broken down into more specific strategies that can be adjusted to suit specific texts, tasks, and situations.

Since the emergence of reading on digital devices, many studies have attempted to determine the similarities and differences between traditional paper-based and digital reading (Afflerbach & Cho, 2010; Coiro & Dobler, 2006; Leu et al., 2015).

Afflerbach and Cho (2010) synthesized the results of 47 studies focusing on reading on the Internet, using read-aloud protocols (Afflerbach, 2000) on one hand, and research into reading multiple or intertextual texts on the other (Rouet & Britt, 2011). According to Afflerbach and Cho (2010), strategies used by expert readers are rather similar for reading texts in print or on the Internet. However, "the Internet represents a change in the architecture of reading" (Afflerbach & Cho, 2010, p. 217). The text to be read is not given, but constructed by the reader's path through pages they decide to visit. A strategy specific to digital reading is therefore, according to Afflerbach and Cho, to "realize and construct the potential texts to be read" (p. 217). Two other strategies used in traditional reading also have greater importance when reading on the Internet: (a) Self-regulation

3. Progress in International Reading Literacy Study see https ://timssandpirls.bc.edu/

or monitoring of the reading act is required given the nebulous, infinite nature of the Internet. Readers need to stay focused on their reading goals and avoid wandering down related paths. (b) Evaluation of the quality of the information or the credibility of the author is also required in traditional paper reading. However, given the unfiltered and often unreferenced information available on the Internet, critical evaluation regarding the credibility or reliability of sources is even more important for digital reading.

Apart from a few differences, there is relative consensus on what distinguishes print reading from digital reading and what makes reading on digital devices particularly complex: (a) text on digital devices has no defined boundaries forcing the reader to construct the text to be read; (b) the reader has to deal with multiple texts and their potential contradictions; (c) the increased importance of self-regulation to focus on reading goals and avoid getting lost in the web; and finally, (d) the crucial role of assessing the credibility of sources. As Li (2020) summarizes, the term novelty referred to in digital reading strategies has two different meanings.

> Novelty about online reading strategies has a twofold meaning. First, it refers to the electronic literacy skills online readers adopt to suit the new online reading environment. Strategies of this kind are new and distinct from the traditional text reading strategies. They are featured by Internet text characteristics and play their unique roles in various Internet reading tasks. (...) Second, novelty refers to the new functions of the traditional strategies transferred to the new online reading contexts. That is, online readers may adapt their reading strategies transferred from traditional reading to suit the new reading contexts (Li, 2020, p. 4).

PISA 2018 in the Era of Digital Reading

Since the reading assessment framework was developed for PISA 2000, the place and functions of reading in society, the devices used and the types of texts, have undergone considerable change. Impacted by information and communication technologies, the way people read, and exchange information has changed rapidly. In 2018, the reading framework underwent substantial revision. Online reading is now central to assessment, to such an extent that all the new assessment tasks developed for 2018 are online reading units[4], including features specific to electronic texts - browsing tools for navigating between pages and hypertext links (Coiro & Dobler, 2006; Leu et al., 2015; OECD, 2019; Rouet, 2006; Rouet & Britt,

4. A reading unit consists of one or more texts, followed by questions.

2011; Rouet & Coutelet, 2008). The three main reading processes assessed in PISA were the same in 2009 and 2018: locating information, interpreting, reflecting, and evaluating texts. However, sub-processes specific to digital reading have been added: *Searching for and selecting relevant texts and Detecting and managing contradictions*.

The contextual questionnaires were updated in a similar fashion, to prominently feature characteristics specific to digital reading alongside aspects of traditional paper reading. To take just two examples, students were asked about their reading practices on paper, but also on digital devices. Two scales measuring metacognitive knowledge about reading already featured in 2009: *Understanding and memorizing a text* (UNDREM) and *Summarizing* (METASUM) (see Appendix 2). In 2018, a third scale *Assessing the credibility of a message* (METASPAM) aims to specifically assess this knowledge about digital reading[5].

Metacognition, Strategies, and Engagement in Reading Tasks

Since the seminal work of Flavell (1976), metacognition has been defined as "any knowledge or cognitive activity that focuses on or regulates an aspect of cognitive activity" (Flavell et al., 2002, p. 150). Theoretical models of metacognition distinguish *metacognitive knowledge* from *metacognitive activities* of controlling or regulating one's understanding (Flavell, 1976; Paris et al., 1983). Numerous studies carried out since the 1980s (Paris & Winograd, 1990; Snow et al., 1998) have highlighted the fact that expert readers differ from less skilled readers with more in-depth metacognitive knowledge and regular use of strategies to regulate their comprehension. In contrast, less skilled readers do not regulate their comprehension and are unaware of alternative strategies to use when they notice a break in comprehension (Paris et al., 1983). Furthermore, work in the field of socioemotional factors related to reading suggests that attitudes, motivation toward reading, self-concept as a reader, and reading self-efficacy can have an impact on how readers mobilize their reading strategies and engage cognitively with the task (Guthrie & Alvermann, 1999; Guthrie et al., 2013; Horner & Shewry, 2002; Lihong et al., 2021; Mc Elwany & Schwabe, 2019).

5. This new scenario was developed by D. Lafontaine, S. Géron and P. Schillings (Université de Liège) as part of the consortium in charge of developing contextual questionnaires for PISA 2018 led by the *Deutsche Institut für Pädagogische Forschung* (DIPF).

Metacognition and Reading Comprehension

The link between metacognition and reading skills has been solidly established by numerous studies focusing on knowledge and use of reading strategies, both quantitative (Baker & Brown, 1984; Denton et al., 2015; Guthrie, et al., 2013) and qualitative using think-aloud protocols (Coiro & Dobler, 2006). High-performing or expert readers have strong knowledge of effective reading strategies. On the other hand, struggling readers know little about them. According to Artelt and Schneider (2015) and Samuelsen and Braten (2007), metacognitive strategies predict reading comprehension more effectively when they focus on a specific reading task.

Ohtani and Hisakawa (2018) have conducted a meta-analysis focusing on the links between metacognition and performance in different disciplines. This meta-analysis of 118 studies includes, among possible moderators, the type of metacognition measure used. On average, metacognition correlated positively ($r = 0.28$) with academic performance. When metacognition is measured via online measures, i.e., think-aloud protocols or log-file data analysis, which records people's behavior while solving a task on digital devices, correlations with performance are higher ($r = 0.53$). In contrast, when metacognition is measured independently, particularly using questionnaires, the correlation with performance is much lower ($r = 0.19$).

Burin et al. (2020) studied the links between metacognition and digital text comprehension in 219 higher education students. To measure metacognition, they used a subset of the Metacognitive Awareness of Reading Strategies Inventory (MARSI) developed by Mokhtari and Reichard (2002), described in the next section. The reading comprehension test comprised 20 items based on two informative texts. The authors also took into account verbal ability and working memory, as well as students' Internet experience. The results of correlation and regression analyses showed that metacognition is significantly correlated with reading comprehension scores ($r = 0.21$), even after controlling for other variables. Verbal ability was also significantly related to reading scores ($r = 0.41$), as was Internet experience ($r = 0.25$). However, Internet experience was not significantly related to the metacognition measure.

Measuring Metacognition in Reading

As an unobservable process, metacognition is difficult to measure (Allen & Armour-Thomas, 1993). According to Mokhtari and Reichard (2002), who surveyed and critically analyzed metacognitive reading

knowledge and activity scales developed before 2000, "efforts to develop inventories of metacognitive knowledge are well-intentioned, but they are generally unsatisfactory from a measurement standpoint" (Mokhtari & Reichard, 2002, p. 250). Their Metacognitive Awareness of Reading Strategies Inventory (MARSI) (Mokhtari and Reichard, 2002) contains 30 items measuring metacognitive knowledge. However, after close examination, the inventory focuses on the perceived use of different reading strategies. The scale, validated with 825 secondary school students (grades 6 to 12), has good reliability ($\alpha = 0.89$) and three factors have been identified: global reading strategies (e.g. I have a goal in mind when I read.), problem-solving strategies (e.g. I try to refocus if I lose concentration.), and reading support strategies (e.g. I take notes when I read.). A systematic survey of three databases (ERIC, Web of science, and APA) showed that MARSI is currently the main instrument used to measure reading metacognition and has been the subject of various adaptations and validation studies (Anderson, 2003; Wu et al., 2012). Anderson (2003) developed an online version of MARSI, but the only change made was the addition of the words: *when I read online* to all 30 items, which are general reading strategies. This does not suffice to make the strategies specific to online reading. In 2018, Mokhtari et al. conducted a validation process of a shorter revised version (15 items) of MARSI (MARSI-R), with a sample of 1,162 high school students (grades 6 to 12). The results confirmed the original three-factor latent structure (see above, Mokhtari & Richard, 2002). MARSI-R showed a moderate correlation of 0.33 with reading self-concept[6] and weak correlations with students' grades, close to zero. Only the *Global reading strategies* correlation ($r = 0.08$) with the scores was significant at $p < 0.05$. Mokhtari et al. (2018) point out:

> One of the persistent problems with MARSI is that correlations are relatively low between reported strategy use scores and external measures of reading performance. (...) Undoubtedly, the problems raised by self-reported instru-

6. "*The variable reader, which asks students to estimate their level of reading ability, comes from the General information section of the MARSI-R and represents respondents' answer to the item: "I consider myself: (1) an excellent reader, (2) a good reader, (3) an average reader, or (4) a poor reader". We found the correlation coefficients, all statistically significant (p <.001), to be (1) r = 0.32 between reader and global reading strategies, (2) r =.346 between reader and problem-solving strategies, (3) r = 0.16 between reader and support reading strategies, and (4) r = 0.33 between reader and the total scale score on the MARSI-R.*" (Mokhtari et al., 2018, p. 235).

ments, concerning generalized vs. contextualized strategy use, play a role in this correlation problem (Mokhtari, et al., 2018, p. 238).

Finally, Li (2020) developed and validated a Second Language Online Reading Strategies Inventory (SLORSI) for ESL with a sample of 482 students from seven Chinese universities. To our knowledge, this is one of the few instruments to have undergone a validation process that actually focuses on online reading strategies. It comprises 29 items in the form of 5-point Likert scales, from *Strongly disagree with this statement* to *Strongly agree with this statement* and is a self-reported measure of metacognitive activities. Factor analyses revealed the existence of nine different factors relating either to traditional reading (inferring, skimming) or online reading (synthesizing, saving, navigating). For our purposes, it is important to note that *evaluating* is a distinct dimension, as it is the focus of our validation study.

The scales just mentioned, regardless of their qualities, are all *self-reported measures* of the respondent's use of different strategies, and concern metacognitive *activities*, not metacognitive *knowledge*. Burin et al. (2020) reiterate the importance of this distinction between metacognitive activities and knowledge. They point out that the only example of metacognitive *knowledge* measure was developed for PISA 2009 with the metacognitive scenarios described in the next section.

Metacognitive Knowledge in PISA 2009 and PISA 2018

A measure of metacognitive *knowledge* was included in the PISA student questionnaire for the first time in 2009 (Artelt & Schneider, 2015; OECD, 2012). The developers adopted an original approach with two particularities. Firstly, it included a reading scenario, assigning a goal for the reading task, thus measuring conditional knowledge that is genuinely task related. The aim of the first scenario is to read a text for understanding and remembering (UNDREM). The aim of the second scenario is to summarize a text (METASUM)[7]. Students are presented with a list of five or six strategies for each scenario, and asked to indicate how effective the strategies are on a 6-point Likert scale, ranging from *Not at all effective* to *Very effective*. Students are not asked to indicate whether they use the various strategies themselves (which would be a self-reported measure), but rather whether they think these strategies are effective in a situated reading

7. The scenarios are presented in Appendix 1 (questions ST164 and 165).

task. Knowledge, or awareness, is thus quite distinct from activity (control or regulation of comprehension), which is rarely, if ever, the case in existing scales. Moreover, the social desirability and acquiescence phenomena inherent in self-reported measures are diminished (He & van de Vijver, 2015) by asking students what they think of the effectiveness of the strategies rather than which strategy they would choose in a similar situation.

In addition, a particular coding method was applied. Rather than considering a priori some answers as correct and others not, the students' answers were compared with those of a panel of reading experts from the countries participating in PISA. This provided an external reference point or benchmark. First, the responses of 64 experts to the same metacognitive scenarios were collected. Then, the experts' responses to all pairs of strategies within a scenario were compared, to retain only those cases where one strategy was clearly judged to be more effective than another. Only pairs of strategies agreed upon by a 75% majority of experts were retained, i.e. 23 pairs in all (nine for the first scenario, eight for the second, and six for the third introduced in 2018, presented later). When the student's rating for each pair matched the experts' ratings, the student scored 1; if not, the student scored 0.

The scores were then added up for each scenario. As a result, students whose judgement of the effectiveness of strategy pairs perfectly matched the experts' judgement received a total score of 23 for all three scenarios. For more details about the coding system, refer to Artelt and Schneider (2015) and Zhou et al. (2020). Artelt and Schneider (2015) carried out a validation study of the two metacognitive scenarios from PISA 2009. They tested the links between metacognitive reading scores and reading performance in PISA 2009 and found high correlations with reading performance, averaging 0.48 across OECD countries.

Purpose of This Study

The aim of the present study is to validate the new scale measuring metacognitive knowledge in digital reading mode developed for PISA 2018. More specifically, the aim is to examine how the PISA 2018 scale compares with other reading measures, particularly performance and various socioaffective scales. The other objective is to examine whether the scale diverges from other constructs relating to information and communication technologies. The boundaries between digital reading and other

related constructs such as digital literacy[8] are tenuous. In addition, as one expert in the field points out, "it's a subject on which the terminology is very confused" (Bawden, 2008, p. 24). For our purposes, we retained Martin's (2006) definition of digital literacy as "the awareness, attitude and ability of individuals to appropriately use digital tools and facilities to identify, access, manage, integrate, evaluate, analyze, and synthesize digital resources, construct new knowledge, create media expressions, and communicate with others. (...)" (p. 154).

According to this definition, the concept of digital literacy is broader than digital reading and overarches it completely. Unlike digital reading, digital or ITC literacy includes objects other than written text (digital resources including images, sound, video, etc.), and places greater emphasis on the technological component and the ability to use digital tools.

As the object of this study is a completely new scale, no previous studies about it or similar scales were available. We were therefore unable to formulate research hypotheses in the strict sense of the term. However, we posed working hypotheses which were tested successively in our analyses. A priori, the scale has three components: *metacognitive knowledge*, *reading*, and *technology* (e-mail message).

Working hypothesis

The working hypotheses were designed to match the step-by-step reasoning used to prove the scale measures what it claims to measure, i.e. (a) metacognitive knowledge; (b) about digital reading, and (c) loosely linked to technology (digital device).

- Working Hypothesis 1: If the METASPAM scale measures *metacognitive knowledge* of reading *strategies*, it should correlate strongly with the other two metacognitive reading scenarios already validated UNDREM and METASUM.
- Working Hypothesis 2: If the METASPAM scale measures *reading* strategies, it should correlate as strongly with reading performance as UNDREM and METASUM.

8. *International Computer and Information Literacy Study. "This type of literacy refers to students' ability to use computers to investigate, create, and communicate in order to participate effectively at home, at school, in the workplace, and in the community*. https://www.iea.nl/studies/iea/icils

– Working Hypothesis 3: If METASPAM measures knowledge of *reading* strategies, it should correlate as strongly with the other non-cognitive variables for reading (interest, self-concept, perceived efficacy) as UNDREM and METASUM.

– Working Hypothesis 4: If METASPAM measures *technological skills*, it should correlate more strongly with non-cognitive ICT-related variables (ICT use and interest and perceived ICT effectiveness) than the other two metacognitive scenarios.

Research Hypotheses

The literature review on the links between metacognition, reading performance, and other variables leads to the following research hypotheses which, though not directly related to the validation approach herein, contribute to the advancement of knowledge on the links between metacognition and digital reading:

– Research Hypothesis 1a: According to the meta-analysis by Ohtani and Hisasaka (2018), the correlation of the metacognitive scenario with reading performance should be relatively low (around 0.20), as the measurement of metacognitive knowledge is carried out via a questionnaire, rather than online.

– Research Hypothesis 1b: According to Artelt and Schneider (2015), Burin et al. (2020) and Schellings et al. (2013), given the scenario is linked to a *contextualized* reading task, the correlation with reading performance could be relatively high.

– Research Hypothesis 2: According to Azevedo et al. (2013), Baker and Wigfield (1999), Guthrie and Alvermann (1999), Horner and Shewry, (2002), the metacognitive scenario should strongly correlate with reading-related socioaffective variables - interest, self-concept, and self-efficacy.

– Research Hypothesis 3: According to Burin et al. (2020), the metacognitive scenario should only weakly correlate, if at all, with ICT-related variables - frequency of use, interest, and self-efficacy.

Methodology

Sample

The study covers all 37 OECD countries, representing over 11 million students. The 42 non-OECD partner countries were not included in the study to ensure a relatively homogenous set of countries from an economic and cultural point of view. The samples per country contain between 4,000 and 8,000 students, with an average age of 15.8 years. The PISA test and questionnaire were completed by a representative sample of the population of 15-year-olds in each country. Samples are drawn according to precise rules and must comply with the standards defined by PISA for results to be published.

The samples are drawn in two stages and stratified proportional to school size. In the first stage, schools are selected (at least 150 schools per country) according to defined criteria (e.g. geographical location, type of school, educational network). In the second stage, 42 students are selected at random from the list of 15-year-olds provided by the school. PISA does not select entire classes.

Variables

Cognitive Variables

PISA reading scores are calculated using the Item Response Model (IRM) (OECD, 2020; Fischer & Molenaar, 1995). The procedure enables the estimation of individual plausible values of reading achievement (Von Davier et al., 2009). Each student is assigned 10 plausible values rather than a single score.

The 10 plausible values generated by the Item Response Model (IRM) were used. This score includes assessments on tasks for three reading processes: 25% for *locate information*, 45% *understand*, and 30% for *reflect and evaluate*.

Metacognitive Variables

Three metacognitive scenarios (see Appendix 1): the two metacognitive scenarios included in the questionnaire since 2009, namely the *Understand and remember* scenario (UNDREM) and the *Summarize* scenario (METASUM), as well as the new metacognitive scenario linked to digital reading (METASPAM). In the latter scenario, students were asked which strategies would be appropriate if they received an e-mail

from their cell phone operator with an attachment announcing that they had won a smartphone. The items are: (a) delete the e-mail, (b) check the sender's address, (c) check the operator's website, (d) open and complete the attachment, (e) reply to the e-mail, and (f) ask for more information.

Non-Cognitive Variables Related to Reading

Four scales measuring non-cognitive aspects of reading, using 4-point Likert scales (*Strongly disagree* to *Strongly agree*): a scale measuring interest in reading (JOYREAD, five items), a scale measuring self-concept in reading with three positively oriented items (SCREADCOMP), a second scale measuring self-concept in reading with three negatively oriented items (SCREADDIFF), and a PISA perceived efficacy scale with three negatively oriented items (PISADIFF). These scales are presented in Appendix 2.

Non-Cognitive Variables Linked to Information and Communication Technologies (ICT)

Three scales measuring practices, attitudes, and motivation related to information and communication technologies (ICT): a 5-point Likert scale (*Never* to *Every day*) measuring frequency of use of digital devices (ENTUSE), a scale measuring interest in them (INTICT), and a 4-point scale of self-efficacy in ICT (COMPICT) (*Strongly disagree* to *Strongly agree*). These scales are presented in Appendix 3.

Data Analysis

Many validation studies rely on factor analysis or latent factor modeling using Mplus. As Loye (2018) argues, validation covers a broad spectrum of approaches and cannot be reduced to a series of technical procedures. She deplores the fact that "The assimilation (*editor's note*: between validation and statistical techniques) is such that sometimes the validation approach is confused with statistical or psychometric techniques, without further ado" (Loye, 2018, p. 101). Neither factorial analysis nor Cronbach's alpha could be applied in the present study due to the particularities of the coding method described in detail in the section presenting the scales for measuring metacognitive knowledge in PISA. As pointed out previously, metacognitive scenario scores result from the comparison of responses to pairs of items provided by two different sets of subjects, the students tested in PISA, and a panel of experts.

As the aim of the present study is to establish whether the PISA 2018 metacognitive scale measures reading knowledge rather than digital literacy knowledge, correlations were used; it is indeed the standard procedure when estimating the predictive value of a measure (Messick, 1990). The variables used are presented in Table 1.

Table 1
Presentation of Variables Used in Correlation Analyses

Metacognitive variables	Scenario Understand and remember (UNDREM)	Scenario Summarize (METASUM)	Scenario Assess credibility (METASPAM)	
Cognitive variables	Reading performance in the PISA test			
Non-cognitive variables related to reading	Interest in reading (JOYREAD)	Self-concept (perception of competence) (SCREADCOMP)	Self-concept (perception of difficulties) (SCREADDIFF)	Perception of the PISA test difficulty (PISADIFF)
Non-cognitive variables related to ICT	Frequency of ICT use (ENTUSE)	Interest in ICT (INTIC)	Interest in ICT (INTIC) ICT self-efficacy (COMPICT)	

To obtain unbiased estimates, correlations took into account specific database characteristics such as sampling mode, replication weights, and the estimation of plausible values (OECD, 2020). IBM SPSS Statistics for Windows (version 26.0) was used for these analyses.

Results

Correlations were first calculated between the METASPAM scenario and the other two metacognitive scales UNDREM and METASUM (Working Hypothesis 1) to detect traces of convergent validity (Campbell & Fiske, 1959). Correlations were then calculated between the three metacognitive scales, reading performance (10 plausible values) on the one hand

(Working Hypothesis 2), and the other non-cognitive variables related to reading (interest in reading, self-concept in reading, and self-efficacy in the PISA test) on the other hand (Working Hypothesis 3). Finally, to detect traces of discriminant validity (Campbell & Fiske, 1959), the three metacognitive scales were correlated with different ICT-related variables (use of and interest in digital tools, sense of digital efficacy) (Working Hypothesis 4).

Table 2

Pearson's R Correlation Coefficients Between the Three Metacognitive Scenarios (METASPAM, UNDREM, and METASUM) and Overall Reading Score (Working Hypotheses 1 and 2)

	METASPAM	**UNDREM**	**METASUM**	**Overall reading score**
METASPAM	1	0.32***	0.38***	0.48***
UNDREM		1	0.47***	0.35***
METASUM			1	0.41***

Legend: *** p.< 0.0001. n = 11,701,146 students

The *assess credibility* scenario (METASPAM) correlates at 0.32 with the *understand and remember* scenario (UNDREM) and at 0.38 with the *summarize* scenario (METASUM). This relatively strong correlation indicates that the METASPAM scenario measures some of the same traits as the two existing scenarios. Working Hypothesis 1 is thus confirmed. The correlation between METASPAM and the two former scenarios is, however, somewhat weaker than the correlation between the two former metacognitive scenarios, suggesting that METASPAM also measures other aspects, which was the aim. The 2018 scenario targets metacognitive strategies assessing the credibility of a message in a digital reading context, whereas the two 2009 scenarios focused on knowledge of metacognitive strategies about reading in general.

Furthermore, the METASPAM scenario correlated at least as well as the existing UNDREM (r=0.35) and METASUM (r=0.41) scenarios with the overall reading score (r=0.48). This result confirms Working Hypothesis 2. This robust correlation of 0.48 is a first indication that the METASPAM scenario measures knowledge relating to digital reading, and not only technological knowledge or digital literacy. The fact that the PISA 2018 reading test focuses on digital reading and is administered

electronically, may explain why METASPAM has a stronger correlation with reading performance than the other two scenarios, which address metacognitive knowledge of reading strategies in general.

Table 3

Pearson's R Correlations Between the Three Metacognitive Scenarios (METASPAM, UNDREM, and METASUM) and Non-cognitive Reading Variables related to reading (Interest, Self-concept, Self-efficacy) (Working Hypothesis 3)

	Interest in reading	**Self-concept (skills)**	**Self-concept (difficulties)**	**Self-efficacy**
METASPAM	0.16***	0.17***	- 0.15***	- 0.24***
UNDREM	0.17***	0.11***	- 0.12***	- 0.18***
METASUM	0.19***	0.16***	- 0.16***	- 0.22***

Legend: *** p.< 0.0001. n = 11,862,220 students

Low to moderate significant correlations (between 0.11 and -0.24) were observed between the three metacognitive scenarios and non-cognitive variables related to reading. The correlations were positive with interest in reading and self-concept (perception of one's competence as a reader). The more students express an interest in reading, the more they perceive themselves as competent readers and the better their metacognitive knowledge. Correlations between the scenarios and self-concept on one hand, self-efficacy in the PISA test, on the other hand, were negative. This was expected, given both scales are negatively oriented. The more reading difficulties students reported, the more they struggled in the PISA test and the worse their metacognitive knowledge.

It is important that the validation process examines whether the METASPAM scenario correlates as well as the two pre-existing scenarios with interest, the two facets of self-concept, and self-efficacy in reading. Examination of the correlations shows this is the case. It even tends to be more so for self-concept and self-efficacy than the other two metacognitive scenarios. This provides further support for Working Hypothesis 3 which stated that the METASPAM scenario does, in fact, measure metacognitive knowledge related to reading. If the METASPAM scenario had shown to have a lower correlation with the socioaffective aspects of reading than the two pre-existing scenarios, it may indicate it measured aspects other than knowledge in reading strategies.

Tablea 4

Pearson's R Correlation Coefficients Between the Three Metacognitive Scenarios (UNDREM, METASUM, and METASPAM) and Non-cognitive ICT Variables (ICT Use and Interest, Perceived ICT Effectiveness) (Working Hypothesis 4)

	Frequency of out-of-school ICT use	Interest in ICT	ICT self-efficacy
METASPAM	- 0.03***	0.07***	0.03***
UNDREM	0.07***	0.04***	- 0.001*
METASUM	0.03***	0.06***	0.02***

Legend: *** p.< 0.0001. * p. < 0.05. n = 10,413,173 students. n is lower because the ICT questionnaire is optional and some countries do not participate.

Although statistically significant, very low correlation coefficients (all below 0.10) were observed between the three metacognitive scenarios and the non-cognitive ICT variables. It is important to point out that in a large-scale survey such as PISA where sample sizes are huge, very low correlation coefficients are nevertheless statistically significant. The value of the coefficient should therefore be used to distinguish correlations that have *pedagogical* significance (at least close to or greater than 0.20) from those close to zero, which reach the threshold of statistical significance due to the huge sample size.

For the present validation approach, it is interesting to verify whether the METASPAM scenario, which includes a digital component with the treatment of an apparently suspicious e-mail, has a higher correlation than the other two scenarios with ICT-related variables. This was not the case. METASPAM does not have a stronger correlation with ICT variables than the other two scenarios. This tends to confirm that ICT literacy plays a minor role in the METASPAM scenario, whereas metacognitive knowledge related to the reading process, i.e. *Assess the credibility of information*, plays a major role. This confirms Working Hypothesis 4.

Discussion

The correlation analyses carried out to validate the scenario aimed at measuring metacognitive knowledge about digital reading produced results that tend to show that this scale in fact measure metacognitive knowledge related to digital reading, rather than digital or technological literacy with no link to reading. The working hypotheses that guided

the validation process show that the METASPAM scenario correlated strongly with the two existing metacognitive reading scenarios already validated (Artelt & Schneider, 2015; Zhou et al., 2020) and with the reading performance of 15-year-olds. Working Hypotheses 1 and 2 are therefore confirmed. Examination of the correlations of the three scenarios with important non-cognitive variables related to reading (interest, self-concept, self-efficacy) shows the absence of notable differences between the new scenario and the two existing metacognitive scenarios. This confirms Working Hypothesis 3 and provides further evidence that the *Evaluate the credibility of a message* scenario is related to reading. Finally, for discriminant validity, the near-zero correlations of the *Evaluate the credibility of a message scenario* with ICT-related variables (use of and interest in ICT, ICT self-efficacy) show that students' knowledge about the suitability of different strategies when faced with a suspicious e-mail is not related to either their familiarity or their self-efficacy with ICT. This confirms Working Hypothesis 4. The scale measures metacognitive knowledge related to digital reading rather than to ICT literacy.

Apart from the validation process, the results of the correlation analyses can be analyzed in the light of the research hypotheses posed. The strong correlation of the *Evaluate the credibility of information* scenario with reading performance (0.48) is higher than is usually the case between metacognitive measures collected by questionnaire with performance. These results are close to those reported by Artelt et Schneider (2015), Burin et al. (2020) and Schellings et al. (2013) and diverge from the results of the meta-analysis by Ohtani and Hisasaka (2018), which pointed to the average weakness ($r = 0.19$) of correlations between performance and metacognitive data collected by questionnaires. Our results therefore confirm Research Hypothesis 1b and reject Research Hypothesis 1a of the present study. This result is probably due to the fact that the scenario developed in 2018 is linked to a contextualized reading task, unlike the decontextualized metacognitive measures of most strategy inventories. Mokhtari and Reichard (2002) had already hypothesized that the low correlation of their MARSI inventory with reading performance might result from the fact that it is a self-reported measure unrelated to a specific reading task. In their validation study of the two PISA 2009 scenarios, Artelt and Schneider (2015) also obtained a robust correlation of 0.48 between the two scenarios and PISA reading scores. The three 2018 scenarios, built on the same model, prove to have stronger links with reading

performance than self-reported measures of strategy use because they combine two advantageous properties. The scenarios include a situational setting that assigns a specific reading goal to the reading task. Also, asking students to express how appropriate the strategies are for the task rather than their self-reported practices, biases associated with self-reported measures (social desirability and tendency to acquiesce) are avoided. This type of contextualized scenario proved to be an interesting alternative when online measures are impossible, as it has the same strong correlation with reading performance as online measures according to Ohtani and Hisasaka's (2020) meta-analysis. Schellings et al. (2013) also showed that when a questionnaire measuring metacognition focuses on a specific reading task - in their case, reading and memorizing a text - the correlations are significantly stronger with reading strategies collected by read-aloud protocols than using more general questionnaires.

As expected, and in line with Research Hypothesis 2, the metacognitive scenario *Evaluate the credibility of a message* as well as the scenarios *Understand and remember* and *Summarize* are significantly correlated with socioaffective variables related to reading (interest, self-concept, self-efficacy), as shown in work by Azevedo et al. (2013), Baker and Wigfield (1999), Guthrie and Alvermann (1999), Horner and Shewry (2002), Mc Elwany and Schwabe (2019) on the links between motivation, engagement in reading, and the implementation of metacognitive and self-regulatory strategies.

Finally, consistent with findings by Burin et al. (2020), the metacognitive scenario is very weakly correlated with ICT-related variables (frequency of use, interest, self-efficacy). Research Hypothesis 3 is therefore also supported.

Conclusion

This study collected sufficient evidence of validity to conclude that the scale assesses metacognitive knowledge about digital reading. The scale correlates strongly with the two scenarios measuring metacognitive knowledge about reading that have already been validated in previous studies (Artelt & Schneider, 2018; Zhou et al., 2020), with reading performance and with socioaffective variables related to reading. We conclude that this scale, aimed at measuring metacognitive knowledge, and more specifically the *Assess the credibility of information* process, is therefore

an interesting tool for quantitative studies about digital reading, particularly given the scarcity of metacognitive knowledge measurement scales. In the completely different context of classroom practice, this scale may help teachers to assess their students' metacognitive knowledge before and after lessons focusing on effective digital reading strategies. It could also be used in a discussion where students are asked to compare the strategies they consider appropriate for dealing with the type of fraudulent messages that regularly clutter our e-mails.

Limits

The scale focuses on a single process, *Evaluate the credibility of information*, whose importance is recognized in a digital reading context. However, this scale alone cannot claim to represent all the processes or strategies involved in digital reading. As a reminder, in the study devoted to an inventory of online reading strategies, Li (2020) identified no fewer than nine factors, including evaluation, alongside other strategies specific to digital reading (information location, synthesis, saving, and browsing). Future studies need to develop other scenarios to evaluate a wider range of digital reading strategies.

Our study covered the 37 OECD countries. The results obtained therefore relate to OECD countries, and cannot be generalized without caution to a wider, more culturally diverse area. Other studies have clearly established the importance of testing the stability or cross-cultural invariance of instruments used in international surveys, and pointed to biases linked to cultural-style responses that were more marked in southern and non-western countries (He & van de Vijver, 2013; Lafontaine et al., 2019). This essential issue was not addressed in the present study.

Finally, the correlations observed between scenarios and reading performance confirm the strong correlations of metacognitive knowledge and reading (Artelt & Schneider, 2015). However, PISA is a cross-sectional study and therefore not an adequate design to draw conclusions about causality. To that end, a longitudinal or quasi-experimental approach would be more appropriate.

Acceptance: January 11, 2022
Final version: June 28, 2022
Acceptance: September 21, 2022

LIST OF REFERENCES

Afflerbach, P. (2000). Verbal reports and protocol analysis. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (volume 3) (pp. 163-181). Longman.

Afflerbach, P., & Cho, B. Y. (2010). Determining and describing reading strategies. Internet and traditional forms of reading. In H. S. Waters & Schneider, W. (Eds.), *Metacognition, use, and instruction* (pp. 201-225). Guilford Press.

Allen, B. A., & Armour-Thomas, E. (1993). Construct validation of metacognition. *The Journal of Psychology: Interdisciplinary and Applied*, *127*(2), 203-211. https://doi.org/10.1080/00223980.1993.9915555

Anderson, N. J. (2003). Scrolling, clicking, and reading English: Online reading strategies in a second/foreign language. *The Reading Matrix*, *3*(3), 1-33.

Artelt, C., & Schneider, W. (2015). Cross-country generalizability of the role of metacognitive knowledge in students' strategy use and reading competence. *Teachers College Record*, *117*(1), 1-32.

Azevedo, R., Behnagh, R., Duffy, M., Harley, J., & Trevors, G. (2013). Metacognition and self-regulated learning in student-centered leaning environments. In D. Jonassen & S. Land (Ed.), *Theoretical foundations of learning environments* (pp. 171-197). Lawrence Erlbaum.

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson (ed.), *Handbook of reading research* (pp. 353-394). Longman.

Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly*, *34*, 452-477.

Bawden, E. (2008) Origin and concepts of digital literacy. In C. Lankshear & M. Knobel (eds.). *Digital literacies: Concepts, policies and practices*. (p. 17-33). Peter Lang Publishing.

Brozo, W., Sulkunen, S., Shiel, G., Garbe, C. Pandian, A. & Valtin, R. (2014). Reading, gender and engagement. *Journal of adolescent literacy*, *57*(7), 584-593.

Burin, D. I., Gonzalez, F. M., Barreyro, J. P., & Injoque-Ricle, I. (2020). Metacognitive regulation contributes to digital text comprehension in E-learning. *Metacognition and learning*, 15, 391-410.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81-105.

Coiro, J., & Dobler, E. (2006). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly Journal*, *42*(2), 214-250.

Coiro, J., Knobel, M., Lankshear, C., & Leu, D. (2014), Central Issues in New Literacies and New Literacies Research. In J. Coiro, M. Knobel, C. Lankshear & D. Leu (Eds.), H*andbook of Research on New Literacies* (pp. 1-22). Lawrence Erlbaum Associates.

Denton, C.A., Wolters, C. A., York, M., Swanson, E., Kulesz, P., & Francis, D. J. (2015). Adolescents' use of reading comprehension strategies: Differences related to reading proficiency, grade level, and gender. *Learning and Individual Differences*, 37, 81-95. https://doi.org/10.1016/j.lindif.2014.11.016

Fischer, G. H., & Molenaar, I. W. (Eds.). 1995. *Rasch models: Foundations, recent developments, and applications*. Springer.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.),

The nature of intelligence (pp. 231-235). Lawrence Erlbaum Associates.

Flavell, J. H., Miller, P. H., & Miller, S. A. (2002). *Cognitive development* (4th ed.). Pearson Education.

Guthrie, J. T., & Alvermann, D. E. (Eds.) (1999). *Engaged reading: Processes, practices, and policy implications*. Teachers College Press.

Guthrie, J. T., Klauda, S. L., & Ho, A. (2013). Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly*, *48*(1), 9-26.

He, J., & van de Vijver, F. J. R. (2013). Methodological issues in cross-cultural studies in educational psychology. In G. A. D. Liem & A. B. I. Bernardo (eds.), *Advancing cross-cultural perspectives on educational psychology: A festschrift for Dennis McInerney* (pp. 39-56). Information Age Publishing.

Horner, S. L., & Shewry, C. S. (2002). Becoming an engaged, self-regulated reader. *Theory Into Practice*, *41*, 102-109.

Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading: Comprehension and assessment* (pp. 71-92). Lawrence Erlbaum.

Lafontaine, D., Dupont, V., Jaegers, D., & Schillings, P. (2019). Self-concept in reading: Factor structure, cross-cultural invariance and relationships with reading achievement in an international context (PIRLS 2011). *Studies in Educational Evaluation*, *60*, 78-89. https ://doi.org/10.1016/j.stueduc.2018.11.005

Leu, D., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C. & Timbrell, N. (2015). The new literacies of online research and comprehension: Rethinking the reading achievement gap. *Reading Research Quarterly*, *50*(1), 37-59. https://doi.org/10.1002/rrq.85

Li, J. (2020). Development and validation of second language online reading strategy inventory. *Computers and education*, *145*, 103733. https://doi.org/10.1016/j.compedu.2019.103733

Lihong, M., Haifeng, L., & Leifeng, X. (2021). Perceived teacher support, self-concept, enjoyment and achievement in reading: a multilevel mediation model based on PISA 2018. *Learning and individual differences*, *85*, 101947. https://doi.org/10.1016/j.lindif.2020.101947

Loye, N. (2018). What if validation was more than a sequence of technical procedures? *Mesure et évaluation en éducation*, *41*(1), 97-125.

Mc Elvany, N., & Schwabe, F. (2019). Gender gap in reading digitally? Examining the role of motivation and self-concept. Journal for educational research online, 11, https://doi.org/ 10.25656/01:16791

McNamara, D. & J. Magliano (2009), Toward a Comprehensive Model of Comprehension. *The Psychology of Learning and Motivation*, *51*, 297-384.

Martin, A. (2006). A European framework for digital literacy. *Nordic Journal of Digital Literacy*, *1*(2), 151-161.https://doi.org/10.18261/ISSN1891-943X-2006-02-06

Messick, S. (1990). *Validity of test interpretation and use*. Educational Testing Service.

Mokhtari, K., Dimitrov, & Reichard, C. A. (2018). Revising the Metacognitive awareness of reading strategies (MARSI) and testing for factorial invariance. *Studies in Second Language Learning and Teaching*, *8*(2), 219-246.

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, *94*, 249-259.

Ohtani, K., & Hisakawa, T. (2018). Beyond intelligence: a meta-analytic review of the relationship among metacognition, intelligence and academic performance. *Metacognition learning*, *13*, 179-212. https://doi.org/10.1007/s11409-018-9183-8

OECD (2012). *PISA 2009 technical report*. OECD Publishing. https://doi.org/10.1787/9789264167872-en

OECD (2019). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing. https://doi.org/10.1787/b25efab8-en

OECD (2020). *PISA 2018 technical background*. OECD Publishing. https://doi.org/10.1787/1b045c06-en.

OECD (2021). *21st-Century Readers: Developing Literacy Skills in a Digital World.*

OECD Publishing. https://doi.org/10.1787/a83d84cb-en

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, *8*, 293-316.

Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15-51). Lawrence Erlbaum Associates.

Rouet, J. (2006). *The Skills of Document Use: From Text Comprehension to Web-based Learning*. Lawrence Erlbaum Associates.

Rouet, J., & Britt, M. (2011). Relevance Processes in Multiple Document Comprehension. In M. McCrudden, J. Magliano & G. Schraw (Eds.), *Text Relevance and Learning from Text Information Age* (pp. 19-52). Information age publishing.

Rouet, J., & Coutelet, B. (2008). The Acquisition of Document Search Strategies in GradeSchool Students. *Applied Cognitive Psychology*, *22*, 389-406. https://doi.org/10.1002/acp.1415

Samuelsen, M. S., & Braten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *British Journal of Educational Psychology*, *77*(2), 351-378. https://doi.org/10.1348/000709906X106147

Schellings, G., Van Hout, B. H. A. M., Veenman, M. V. J., & Meijer, J. (2013). Assessing metacognitive activities: the in-depth comparison of a task-specific questionnaire with think-aloud protocols. *European Journal of Psychology of Education*, *28*, 963-990. https://doi.org/10.1007/s10212-012-0149-y.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. National Academy Press. http://www.ierinstitute.org/IERI_Monograph_Volume_02_Chapter_01.pdf

Snow, C., & the RAND Corporation (2002). *Reading for Understanding: Toward an R and D Program in Reading Comprehension*. RAND Reading Study Group.

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI monograph series*, 2(1), 9-36.

Wu, L., Valcke, M., & Vankeer, H. (2012). Validation of a Chinese version of the Metacognitive awareness of reading strategies. *Eurasian Journal of Educational Research*, *48*, 117-134.

Zhou, J., He, J., & Lafontaine, D. (2020). Cross-cultural comparability and validity of metacognitive knowledge in reading in PISA 2009: a comparison of two scoring methods. Assessment in Education: *Principles, Policy & Practice*. *27*(6). 635-654. https://doi.org/10.1080/0969594X.2020.1828820

Appendix 1: The Three Scenarios for Assessing Metacognitive Knowledge: Understand and remember (ST164), Summarize (ST165), Assess the credibility of a message (ST166)

ST164

Reading Task: You have to understand and remember the information in a text.

How do you rate the usefulness of the following strategies for understanding and memorizing the text? (Please select one response in each row.)

	Not useful at all (1)	**(2)**	**(3)**	**(4)**	**(5)**	**Very useful (6)**
I concentrate on the parts of the text that are easy to understand.	○	○	○	○	○	○
I read through the text twice quickly.	○	○	○	○	○	○
After reading the text, I discuss its content with other people.	○	○	○	○	○	○
I underline important parts of the text.	○	○	○	○	○	○
I summarize the text in my own words.	○	○	○	○	○	○
I read the text aloud to another person.	○	○	○	○	○	○

ST165

Reading Task: You have just read a long and rather difficult two-page text about fluctuations in the water level of a lake in Africa. You have to write a summary.

How do you rate the usefulness of the following strategies for writing a summary of this two-page text? (Please select one response in each row.)

	Not useful at all (1)	**(2)**	**(3)**	**(4)**	**(5)**	**Very useful (6)**
I write a summary. Then I check that each paragraph is covered in the summary, because the content of each paragraph should be included.	○	○	○	○	○	○
I try to copy out accurately as many sentences as possible.	○	○	○	○	○	○
Before writing the summary, I read the text as many times as possible.	○	○	○	○	○	○
I carefully check whether the most important facts in the text are represented in the summary.	○	○	○	○	○	○
I read through the text, underlining the most important sentences. Then I write them in my own words as a summary.	○	○	○	○	○	○

ST166

Reading Task: You have received a message in your inbox from a well-known mobile phone operator telling you that you are one of the winners of a smartphone. The sender asks you to click on the link to fill out a form with your data so they can send you the smartphone.

In your opinion, how appropriate are the following strategies in reaction to this e-mail? (Please select one response in each row.)

	Not appropriate at all (1)	**(2)**	**(3)**	**(4)**	**(5)**	**Very appropriate (6)**
Answer the e-mail and ask for more information about the smartphone	○	○	○	○	○	○
Check the sender's e-mail address	○	○	○	○	○	○
Click on the link to fill out the form as soon as possible	○	○	○	○	○	○
Delete the e-mail without clicking on the link	○	○	○	○	○	○
Check the website of the mobile phone operator to see whether the smartphone offer is mentioned	○	○	○	○	○	○

Appendix 2: Scales Measuring Non-cognitive Aspects of Reading: Interest in Reading (ST160), Self-concept (ST161), Self-efficacy in the PISA Test (ST163)

ST160

How much do you agree or disagree with these statements about reading?

(Please take into account various kinds of reading material, such as books, magazines, newspapers, websites, blogs, e-mails, etc.) (Please select one response in each row.)

	Strongly disagree	**Disagree**	**Agree**	**Strongly Agree**
I only read if I have to.	○	○	○	○
Reading is one of my favourite hobbies.	○	○	○	○
I like talking about books with other people.	○	○	○	○
For me, reading is a waste of time.	○	○	○	○
I only read to get information I need.	○	○	○	○

ST161

How much do you agree with the following statements? (Please select one response in each row.)

	Strongly disagree	**Disagree**	**Agree**	**Strongly agree**
I am a good reader.	○	○	○	○
I am able to understand difficult texts.	○	○	○	○
I read fluently.	○	○	○	○
I have always had difficulty with reading.	○	○	○	○
I have to read a text several times before completely understanding it.				
I find it difficult to answer questions about a text.	○	○	○	○

ST163

In the PISA test you took before the break, you had to read several texts and answer reading comprehension questions.

How do you feel about these reading tasks? (Please select one response in each row.)

	Strongly disagree	**Disagree**	**Agree**	**Strongly agree**
There were many words I could not understand.	○	○	○	○
Many texts were too difficult for me.	○	○	○	○
I was lost when I had to navigate between different pages.	○	○	○	○

Appendix 3: ICT scales Frequency: Use of ICT Outside School (IC008), Interest in ICT (IC013), Self-Efficacy in ICT (IC014)

How often do you use digital devices for the following activities outside of school? (Please select one response in each row.)

		Never or hardly ever	Once or twice a month	Once or twice a week	Almost every day	Every day
IC008Q01TA	Playing one-player games.	O 01	O 02	O 03	O 04	O 05
IC008Q02TA	Playing collaborative online games.	O 01	O 02	O 03	O 04	O 05
IC008Q03TA	Using email.	O 01	O 02	O 03	O 04	O 05
IC008Q04TA	<Chatting online> (e.g. <MSN®>).	O 01	O 02	O 03	O 04	O 05
IC008Q05TA	Participating in social networks (e.g. <Facebook>, <MySpace>).	O 01	O 02	O 03	O 04	O 05
IC008Q07TA	Playing online games via social Sims Social>).	O 01	O 02	O 03	O 04	O 05
IC008Q08TA	Browsing the Internet for fun (such as watching videos, e.g. <YouTube™>).	O 01	O 02	O 03	O 04	O 05
IC008Q09TA	Reading news on the Internet (e.g. current affairs).	O 01	O 02	O 03	O 04	O 05
IC008Q10TA	Obtaining practical information from the Internet (e.g. locations, dates of events).	O 01	O 02	O 03	O 04	O 05
IC008Q11TA	Downloading music, films, games or software from the internet.	O 01	O 02	O 03	O 04	O 05
IC008Q12TA	Uploading your own created contents for sharing (e.g. music, poetry, videos, computer programs).	O 01	O 02	O 03	O 04	O 05
IC008Q13TA	Downloading new apps on a mobile device.	O 01	O 02	O 03	O 04	O 05

Thinking about your experience with digital media and digital devices: to what extent do you disagree or agree with the following statements?

(Please think of different kinds of digital devices such as for example desktop computers, portable laptops, notebooks, smartphones, tablet computers, cell phones without internet access, game consoles, or internet-connected television) (Please select one response in each row.)

		Strongly disagree	Disagree	Agree	Strongly agree
IC013Q01NA	I forget about time when I'm using digital devices.	O 01	O 02	O 03	O 04
IC013Q04NA	The Internet is a great resource for obtaining information I am interested in (e.g. news, sports, dictionary).	O 01	O 02	O 03	O 04
IC013Q05NA	It is very useful to have social networks on the Internet.	O 01	O 02	O 03	O 04
IC013Q11NA	I am really excited discovering new digital devices or applications.	O 01	O 02	O 03	O 04
IC013Q12NA	I really feel bad if no internet connection is possible.	O 01	O 02	O 03	O 04
IC013Q13NA	I like using digital devices.	O 01	O 02	O 03	O 04

Thinking about your experience with digital media and digital devices: To what extent do you disagree or agree with the following statements?

(Please think of different kinds of digital devices such as for example desktop computers, portable laptops, notebooks, smartphones, tablet computers, cell phones without internet access, game consoles, or internet-connected television)

		Strongly disagree	**Disagree**	**Agree**	**Strongly agree**
IC014Q03NA	I feel comfortable using digital devices that I am less familiar with.	○01	○02	○03	○04
IC014Q04NA	If my friends and relatives want to buy new digital devices or applications, I can give them advice.	○01	○02	○03	○04
IC014Q06NA	I feel comfortable using my digital devices at home.	○01	○02	○03	○04
IC014Q08NA	When I come across problems with digital devices, I think I can solve them.	○01	○02	○03	○04
IC014Q09NA	If my friends and relatives have a problem with digital devices, I can help them.	○01	○02	○03	○04