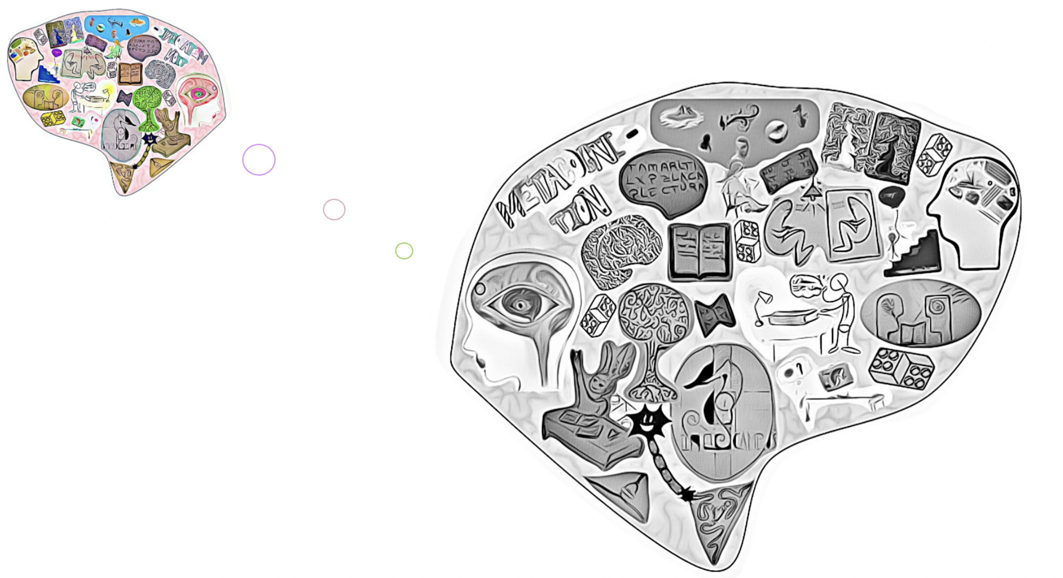
eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

NAZIOARTEKO BIKAINTASUN CAMPUSA
CAMPUS DE EXCELENCIA INTERNACIONAL

The role of metacognitive monitoring in regulating learning in early readers



Doctoral thesis by:
Ioanna Taouki

Supervised by:
Prof. David Soto and **Dr. Marie Lallier**

Donostia/San Sebastián, 2021

The role of metacognitive monitoring in regulating learning in early readers

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

bcbl
BASQUE CENTER ON COGNITION, BRAIN AND LANGUAGE

INPhINIT
"la Caixa" Fellowship Programme

Doctoral thesis by:
Ioanna Taouki

Supervised by:
Prof. David Soto and **Dr. Marie Lallier**

A thesis submitted for the degree of
Doctor of Philosophy in Cognitive Neuroscience

Donostia/San Sebastián, 2021

Basque Center on Cognition, Brain and Language

Paseo Mikeletegi 69

Donostia-San Sebastián, Spain

October 2021

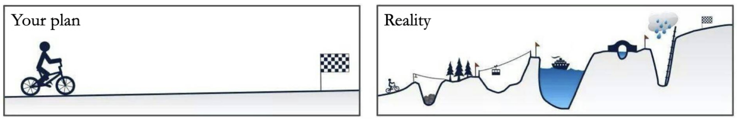
This work received support from “la Caixa” Foundation (ID 100010434) through the fellowship LCF/BQ/DI17/11620070, and the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 714673.

Moreover, support was provided by the same program to conduct a four-month virtual research stay at the MRC Cognition and Brain Sciences Unit, University of Cambridge, under the supervision of Dr. Olaf Hauk. The study carried out during the research stay was not included in the thesis, but the following article is in preparation: Taouki I., Lallier M., Soto D., Hauk O. (2021). “Theta oscillations support metacognitive monitoring in early childhood”.

Acknowledgments

"As you set out for Ithaka, hope the journey is a long one, full of adventure, full of discovery"

(C. P. Cavafy)



Pursuing a Ph.D. is not about receiving a title, nor will you find a linear way to get to the target, but one learns a lot during this full of peaks and valleys process. Not only about science, but also about oneself. During a Ph.D. you get to have "eureka" moments, but you also face failure and you learn how to cope with it. You get to know what teamwork is about, but also how to stand on your own feet when you are the only one studying this specific single piece of the huge puzzle of the mysteries of life. I could write a whole list of experiences and skills that one can develop during this process, but after all, what stays with us, when time passes, are the feelings that the journey brings and our companions in this adventure. So, I would like to dedicate this section to people who accompanied me on this adventure, and hopefully in the next ones as well.

First and foremost, I would like to sincerely thank my main guides in this journey, my supervisors, Prof. David Soto and Dr. Marie Lallier. I want to thank you both equally, for always being there, with the door (or Zoom) open. Your love and enthusiasm for research, combined with prioritizing work-life balance, was a source of admiration for me in life. David, as the expert in metacognition, you taught me how to reflect on each step of the research process, to be a better analytical thinker, and to question everything twice. Marie, as the expert in reading development, you taught me how to organize my thinking and our research ideas, how to improve my academic writing skills, and to convey clear messages to the reader. I speak from the heart when I say that you were the best combination of supervisors I could get for my thesis.

Second, I want to deeply thank the "hidden heroes" contributing to this thesis. Mamen, without you this thesis wouldn't have been real. You were always there, with a big smile, willing to help, even out of your working hours, bringing an optimistic approach to life. I cannot express enough my gratitude to you and the work you did with our kids participants in Carmelitas. Moreover, many thanks to the other groups of BCBLians that make our work possible and our lives easier in the BCBL: the admin, the lab, and the IT team. Thank you Ana Fernandez for helping us with our grants and for your pure kindness, Eider Juaristi, for your willingness to help, your reminders and support with the university procedures, Larraitz Lazcano, Maider Goñi, Vanessa Gallardo, for your management skills. Thank you, Larraitz Lopez and Oihana Vadillo for making our experiments come true and organizing everything with the schools and the lab. Thank you Brendan for prioritizing worker's mental health in the BCBL. And to everyone managing the BCBL, I know you always listen to our complaints, but I think that we all appreciate the warm and supporting environment of the BCBL, and how it develops over the years.

Now, to the people who set out for a similar mission to mine, the PreDoc group. I have to admit that I was surprised to find such a unified team when I first arrived at the BCBL. During these years, I saw every one of this team taking initiatives and supporting each other, in and outside the working place. And mostly, to keep doing so in difficult moments, such as the pandemic. Thank you Piermatteo, Jose, Christoforos, Irene, Candice, Sandra, Teresa, Jose A., Polina, Jordi, Stefano, Mina, Ning, Patxi, Suang, Vicente, Chiara, Catherine, Alberto, Dani, Eneko, Marta, Kshipra, for making me feel part of this team, since day 1, and I hope that the new generations will keep up the sense of unity that we have built.

In each adventure, it's impossible not to have your favorites. I met these girls in the BCBL, but I now consider them friends for life; Maria, you were sitting across my desk at work and you added a lot of humor and moments to remember at the BCBL. And that was only the beginning. Thank you chica for always being there, for making me laugh and see life more positively even in tough moments. Pavlina, I could write a separate thesis on the moments I had with you and your family in Donosti, the BBQs, night swims, the music we played. Most of all, thank you for all the love and homey feeling I have whenever I am surrounded by you, Andrew, little Sophie, Albertito and also the Paus and Kevin, another part of the Donostian family. Dana, I always admired you and your open-mindedness. Thank you for bringing us to earth ("Lur"), and for always smiling. Thank you Maria K. for taking care of me and for bringing the Greek spirit in the BCBL. Last, my Greek teammate, Alexia, the "calm strength". Little did I know, when I first contacted her to orientate me for my moving to Donosti, that we would walk to work together almost every day, having deep conversations about everything. Alexia, I consider you as one of my closest friends and I hope to keep having these moments with you in our next adventures. Thank you also for introducing me to the "postmodern antisocial", but very cool, old Predocs group. Thank you, Jovana and Bojana, the

Serbian dream team, for welcoming me into your group even if I arrived so late. For me, you serve as an example of hospitable and generous people.

Thanks to the "La Caixa" foundation for giving me the opportunity to start this journey in the first place, and for organizing amazing trainings, where we could learn a lot and meet wonderful people. I would especially like to thank Marta. She was the companion with whom I shared most of the peaks and valleys, the joys and the pains, of this journey and the person with the most will I know to climb up the highest peaks of life. Thank you for your "amb energies i illusió", "life is a joke", and "la vida es bonica, pero a vegadas complicada" statements. I hope that you keep bringing this inspiration and love for life to people surrounding you in all of the following life adventures.

To my friends from Greece, even though we were far away during these years, our friendship kept strong and supportive. It is so beautiful to know that some few people in life care about every life adventure of yours and that sooner or later you will meet and share your experiences, and it will feel like no time has passed. Lida, Maria, Elpida, Despina, and Elizaki thank you deeply for that.

The last mentioned, but most important expressed gratitude goes for my family in Greece. Mom, Dad, as I grow older, I realize more and more how complicated a task might be to be a parent. You are always there showing unconditional love, supporting us in each and every step, wherever on the globe we happen to be. It is thanks to you that I had the chance to experience this beautiful journey and I hope you can feel the gratitude and love I have for you. Jason, I feel so lucky to have a brother like you, so caring and supporting. And how could I forget the support of my favourite cousins; Ivi, Mike, Ilias, Antonis, Savina, Roza, Giorgos, and my uncles and aunts. During this thesis, I learned a lot about time management, and I realized that spending time with all of you is the most precious.

And for those of you who say that "No one will read your thesis, apart from your supervisors", my grandma will prove your hypothesis wrong. Giagia, I feel grateful to have a wise grandmother with an innate desire to endlessly learn. I hope to follow your example in life, besides any title and any certificate.

Finally, I want to thank Donosti, a paradise on earth. The landscapes, smells and tastes, and the most colourful, but also the most grey skies I've seen, will always stay in my memory.

Abstract

Metacognition refers to the capacity of reflecting upon our own cognitive processes and entails the abilities of monitoring and regulating our performance in online cognitive tasks. There is an ongoing discussion in the literature on the role of metacognitive monitoring in learning and academic achievement. However, little is known about the developmental trajectories of metacognitive functions during the first years of primary school, namely, when children begin to receive formal education in reading and develop their orthographic lexicon. Moreover, previous literature has highlighted the need of measuring metacognitive monitoring, using indexes which control for various confounding individual factors, such as the overall tendency of participants to be over or under confident, or their overall capacity to perform a task.

In the context of the present thesis, a within-subject longitudinal study was carried out to evaluate students' metacognitive monitoring ability in tasks related to orthographic lexical processing and its prerequisites, and a non-reading related task, during earlier and later stages of reading development (from Grade 1 to Grade 3). A hierarchical Bayesian Signal Detection Theory (SDT) model was used for the estimation of metacognitive ability in each task and time point of the study, in order to avoid the confounding effects of the above-mentioned factors.

The main goals of this study were threefold. First, we aimed to investigate how metacognitive monitoring develops within linguistic and non-linguistic tasks during the first years of primary school. Second, we assessed whether or not young children recruit common or domain-specific mechanisms supporting metacognition across the different task domains and whether this pattern changes between Grade 1 and Grade 3. Third, we sought to examine whether and how metacognitive monitoring ability (i.e., how confidence ratings track accuracy in the task) and task performance in the linguistic and non-linguistic tasks relate to students' standardized reading ability during the first years of primary school. Finally, we asked whether early metacognitive ability in reading-related tasks predicts longitudinal improvements in students' reading performance.

No association was found, in any stage of reading development studied here, between students' metacognition in the reading-related tasks and performance on the standardized reading tests, notwithstanding first-order performance correlated across these tasks. Remarkably, early

metacognitive ability was negatively associated with task performance in tasks assessing orthographic lexical processing in Grade 1 and positively predicted childrens' performance improvement in reading performance two years later. Conversely, early task performance in the same tasks negatively predicted reading performance improvements across time. These findings indicate that students who are less experienced in reading in the beginning of primary school, use metacognition and efficient error-monitoring as a tool to catch up with their more advanced peers.

Against our expectations, we found that signal-detection measures of metacognitive efficiency decreased across time in both the linguistic and the non-linguistic tasks. This result may suggest that students' experience and improvement on the tasks led to a decreased need of metacognitive resources, favouring more automatic reading strategies. In Grade 1, we only found little evidence consistent with domain-general mechanisms supporting metacognition. However, this was not borne out in Grade 3 given that no significant association was observed in students' metacognitive efficiency between any pair of tasks.

Taken together, results of the present study suggest that the development of metacognitive processing may be dissociated to some extent from reading-related linguistic abilities during the first years of primary school. Nevertheless, it may play a fundamental role in guiding students' learning. These data highlight the importance of creating educational programs fostering students' metacognition as a long term learning tool.

Table of contents

Abbreviations and acronyms list

AUROC2	Area under the receiver operating characteristic type-2 curve
DRC	Dual route cascade
FDR	False discovery rate
fMRI	Functional magnetic resonance imaging
FOK	Feelings of knowing
HDI	Highest density interval
JOL	Judgments of learning
MANOVA	Multivariate analysis of variance
Mratio	Metacognitive efficiency
MTM	Multiple trace model of polysyllabic word reading
PDP	Parallel distributed processing model
PFC	Prefrontal cortex
SD	Standard deviation
SDT	Signal detection theory
VA	Visual attention

1. Chapter I: General Introduction

"We do not learn from experience, we learn from reflecting on experience." - John Dewey

"The unexamined life is not worth living." - Socrates

Reflecting on our knowledge has been long considered an essential skill for learning. The concept of thinking about our own experiences, roots back to ancient philosophers, and continues to be investigated up to date under the term "metacognition" (Flavell, 1979; Metcalfe & Shimamura, 1994). The use of the prefix *"meta"* means after or beyond, hence the term metacognition in experimental psychology and cognitive neuroscience currently stands as the study of 'cognition about cognition' or the ability to introspect about one's own cognitive and behavioural processes. During the last decades, an increasing interest has been shown in clarifying the role of metacognition in learning and academic achievement, and in developing intervention strategies in order to improve students' metacognitive abilities in different domains, such as reading and mathematics (Fleur et al., 2021).

Efficient text reading is a complex skill, which is fundamental for our everyday life and for providing us access to a wide range of learning resources. Among other domains, a special interest has been developed lately in understanding how individuals monitor their reading and their comprehension of written text (e.g., Dunlosky & Lipko, 2007; Thiede et al., 2009). However, this body of research has been focused on the study of metacognition in terms of "meta-comprehension" in skilled readers, and the role of metacognition during the early stages of reading development, such as the stage of visual word recognition, has not yet been investigated.

The *present thesis* aims to investigate how metacognition relates to reading acquisition during the first years of primary school.

1.1. Metacognition and metacognitive monitoring

1.1.1. Definition and models of metacognition: metacognitive monitoring and control

If you go back to your school times, I believe you can all recall your teachers using one of the following phrases: "Are you sure that this is the correct answer?", "Have a better think on this", "Learn from your mistakes", "You need to study these concepts more", "Respond to the questions of the exam that you feel more confident in first, and then leave time to devote to the hard questions". These phrases point to one's capacity of monitoring their learning processes and building on top, or modifying already acquired knowledge in order to improve learning, what we now call "metacognition".

Metacognition refers to the ability of an individual to reflect on their own cognition and behaviour, e.g., to track the correctness of one's thoughts, actions, and behavioural responses (Metcalfe & Shimamura, 1994) across multiple task contexts (Narens, 1990), and was first introduced as a term by Flavell (1979), defined as our ability to "think about thinking" (Flavell, 1979).

Over the last decades, metacognition has gained a lot of interest in various fields, such as in educational research, as a candidate predictor of academic achievement and learning (Efklides, 2011), in the field of experimental psychology, related to visual perception, decision making and meta-memory (Koriat & Goldsmith, 1996), in developmental psychology (Lockl & Schneider, 2006; McCurdy et al., 2013) and in clinical settings, related to brain lesions and awareness impairments in dementia patients (Molenberghs et al., 2016; Rouault et al., 2018; Koriat & Goldsmith, 1996).

Due to the different scopes of each research field, the relationship of metacognition with several cognitive functions has been examined, i.e., theory-of-mind, consciousness, self-regulated learning, memory, executive functions, motivation, and social cognition (Efklides, 2011; Koriat & Goldsmith, 1996; Lockl & Schneider, 2006). Several theoretical frameworks for studying metacognition have been put forward during the last decades. For instance, Flavell (1979) classified metacognition into two components; *declarative metacognition*, which refers to an individuals' declarative knowledge about their own cognition, and *procedural metacognition,* which englobes the higher-order cognitive processes taking place when an individual is performing a cognitive task, involving the regulation of her own ongoing performance (Flavell, 1979).

A widely used metacognitive framework in educational settings has been proposed by Efklides (2008, 2011), who stratified metacognition in: a) *metacognitive knowledge* or *metacognitive awareness*, which is an analogue of Flavell's declarative metacognition, referring to knowledge about our own and other people's cognitive processes, b) *metacognitive experiences,* which include knowledge, feelings, and judgments generated in on-line task performance and c) *metacognitive strategie*s/skills, referring to the intentional employment of cognitive strategies in order to regulate cognition and guide behaviour (Efklides, 2008, 2011). In the present study, we will focus on the concept of procedural metacognition/metacognitive experiences component, which relates directly to online processes associated with task performance.

Influential models of procedural metacognition propose that metacognition is mediated by an interaction between an object-level process (e.g., a reading or perceptual task, namely, the *object-level* or *type-1 performance*) and a second-order process (e.g., the *meta-level* or *type-2 performance*, see Figure 1). The meta-level component monitors the first-order process and, when cognition fails (i.e., following an error), exerts control processes in order to promote adaptive behaviour (Koriat & Goldsmith, 1996; Nelson, 1990). These two processes have been referred to as "metacognitive monitoring" and "metacognitive control" respectively.

Nelson and Narens (1994) specifically focused on the function of these metacognitive processes during students' learning. They suggested that upon knowledge acquisition, efficient metacognitive monitoring evolves when students evaluate the recently presented information from the teacher and their level of certainty in understanding and retaining this information. Efficient metacognitive control occurs when students decide to allocate further studying time to the non-well established information received (Nelson et al., 1994).

Metacognitive monitoring is assessed in the lab by collecting subjective judgments, asking individuals to give either *prospective judgments of learning* (*JOLs)* or feelings of knowing (FOKs), by predicting how well they have learned/how well they will remember the information presented respectively (Koriat & Ackerman, 2010; Metcalfe, 2009; Metcalfe & Finn, 2013), or *retrospective confidence judgments*, by assessing their level of certainty regarding their accuracy of responses.

Conversely, metacognitive control is assessed by tracking the time individuals are allocating to study a piece of information (Destan et al., 2014; Metcalfe & Finn, 2013).

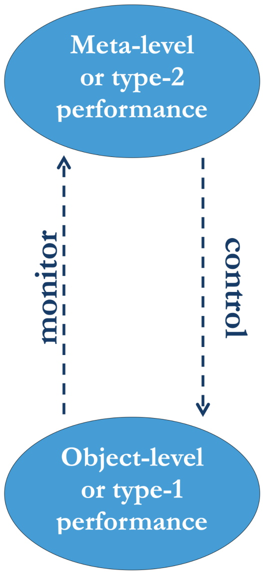


Figure 1 Theoretical mechanism of metacognition (Nelson & Narens, 1994).

Metacognitive monitoring and control have been considered to reciprocally influence each other, with metacognitive monitoring suggested to be a prerequisite of metacognitive control (Metcalfe, 2009), and metacognitive control found to have an impact on future monitoring. For instance, the time it takes individuals to respond in a type-1 task (response latency) has been suggested to negatively relate to participants' confidence judgments (Koriat & Ackerman, 2010).

The present study will focus on assessing the component of metacognitive monitoring by collecting confidence judgments, a tool that is currently being used widely in empirical studies in experimental psychology and cognitive neuroscience (e.g., Baird et al., 2013; Fleming & Lau, 2014; McCurdy et al., 2013).

1.1.2. Assessment of metacognitive monitoring

As mentioned above, a useful tool to measure the efficiency of individuals' explicit metacognitive monitoring ability is the use of retrospective trial-by-trial confidence judgments. Following the first-order decision (type-1) related to the primary task (e.g., discriminating the category of briefly presented sequences of letters representing either words or pseudowords), participants are asked to rate how confident they are about the correctness of the first-order task response (Fleming & Lau, 2014). A metacognitive ideal individual assigns high confidence ratings to correct type-1 responses and low confidence ratings to incorrect type-1 decisions.

The relationship of confidence judgments and task accuracy, or metacognitive resolution/sensitivity, was initially measured using simple correlations (phi and gamma correlations). Phi correlation is calculated by simply calculating the Pearson's r coefficient between the trial-by-trial accuracy (correct vs incorrect response) and participants' confidence judgment (high vs low certainty) (Kornell et al., 2007). Gamma Goodman Kruskal correlations involve a non-parametric measure of metacognition calculated as the difference between correct trials rated with high confidence and incorrect trials rated with high confidence divided by the total number of trials (Aldrich & Nelson, 1984; Koriat et al., 2009). Another method used to estimate the relationship between confidence judgments and task performance is the so-called calibration or absolute monitoring accuracy. This measure quantifies the degree that subjective confidence judgments track objective accuracy in task performance. When the percentage of responses rated with high confidence is higher than the percentage of correct responses (objective accuracy), overconfidence occurs, while in the opposite case, underconfidence occurs (e.g., Brier score, QSR score, for a review see Fleming & Lau, 2014).

All the above-mentioned methods have been criticized because they can be confounded by the individual confidence bias, defined as the tendency of a participant to use higher or lower confidence ratings in a cognitive task (Masson & Rotello, 2009). This is particularly critical when assessing metacognitive function in young children because they typically show an overconfidence bias (Finn & Metcalfe, 2014). To tackle this issue, Galvin (2003) developed a measure of "type 2 sensitivity" based on Signal Detection Theory (SDT) (Green & Swets, 1988; Wickens, 2001), which models metacognition based on the area under the type-2 Receiver Operating Characteristic curves (AUROC2), which can provide an estimate of metacognitive sensitivity independently of the individual confidence bias (Galvin et al., 2003). However, Galvin's model is still affected by the confounding factor of type-1 sensitivity, meaning that participants who perform better in the type-1 task may erroneously appear to also have better metacognitive sensitivity compared to their peers (Fleming & Lau, 2014; Masson & Rotello, 2009).

Recently, Maniscalco and Lau (2012) have developed an SDT confidence bias-free model for calculating metacognitive sensitivity (meta-d'), which is on the same scale and hence comparable with the individual type-1 sensitivity (d-prime or d'). This allows for an estimate of type-2 performance, which controls for participant's type-1 performance, defined as metacognitive efficiency (Mratio or meta-d'/d'). This measure permits meaningful comparisons of metacognitive efficiency across participants or tasks (Maniscalco & Lau, 2012), for a review see (Fleming & Lau, 2014). A more recent Bayesian framework, based on this model, has also optimized the estimates of metacognitive efficiency when handling datasets with few trials coming from patients or children (Fleming, 2017). Hence, this will be the primary measure used in the present study. The use of this model is described in detail in the Chapter Section 1.3.2.4.

1.1.3. Metacognitive monitoring during development

The ability to accurately assign confidence ratings to type-1 performance has been suggested to develop with age, though understanding of its developmental course remains incomplete (Palmer et al., 2014; Weil et al., 2013). Evidence for metacognitive processing has been found early on during development in infants as young as 12-months old showing increased persistence in correct vs incorrect choices in a task, which indicates the existence of an implicit internal monitoring system of one's performance from the very early stages of life (Goupil & Kouider, 2016), while 20-months old infants have shown the capacity of seeking help when they don't know (Goupil et al., 2016).

The specific developmental stage in which children start to accurately provide explicit verbal metacognitive monitoring judgments to track their task performance is still under debate. Few recent studies have suggested that already from the age of 3 children assign higher confidence ratings to correct than incorrect responses in perceptual tasks, but not in memory tasks, when these depend less on verbal reports (Ghetti et al., 2013; Lipowski et al., 2013). Notwithstanding, several research studies have suggested that children's explicit metacognitive abilities are still poor under the age of 4, and especially in conditions in which the student has partial knowledge on the perceptual task performing (Rohwer et al., 2012; Sodian et al., 2012). This has often been attributed to an innate tendency of children in early childhood to show overconfidence and wishful thinking when judging the correctness of their responses, both in experimental cognitive tasks (e.g., Finn & Metcalfe, 2014; Lipko et al., 2012) and in real-life settings (Schneider, 1998), which does not improve by practice on the task or even with explicit feedback (Finn & Metcalfe, 2014; Lipko et al., 2012).

Destan et al. (2014) cross-sectionally assessed the likelihood that children assign low confidence ratings in incorrect and high confidence in correct responses in children of 5, 6, and 7 years old and found no differences between the three age groups. However, after the age of 6 children were shown to make good use of these confidence judgments in order to control their subsequent performance on the task (Destan et al., 2014; Rohwer et al., 2012). The unchangeable monitoring ability in these ages may well be related to the fact that childrens' pronounced overconfidence has been found to decrease between the ages of 7 to 10, which has been specifically linked to their increasing ability to accurately monitor incorrect responses (Schneider, 2015). Indeed, in another study, Krebs et al. (2010) showed that monitoring ability increased significantly between children of 8 and 12 years old in a close response test on a previously presented educational film (Krebs & Roebers, 2010), while Roebers and Spiess (2017) assessed this ability in a spelling task using a longitudinal within-subject design and found that it increased over the course of one school year (8-9 years olds participants) (Roebers & Spiess, 2017).

Based on these findings, one could safely assume that the first years of primary school play a significant role in the development of one's ability to accurately judge their confidence on type-1 decisions. However, it is important to mention that all the above-mentioned studies used metacognitive indexes which are susceptible to type-1 performance and confidence biases. In the present study, a bias-free measure of metacognitive efficiency will be used to avoid these confounding effects. Due to the fact that this measure has been recently developed, there is, to our knowledge, no research investigating the developmental course of metacognitive efficiency during early childhood. Metacognitive efficiency has been studied over the course of adult life (18-84 years old participants) and has been proposed to decrease significantly over the years in a visual perceptual task (Gabor test), and non-significantly in a memory task (Palmer et al., 2014). In the present study, we aim to assess metacognitive efficiency in children during the first years of primary school (6-9 years old).

1.1.4. Neural basis of metacognitive monitoring

Metacognitive monitoring is a higher-order brain process that has been linked to the activity of frontoparietal networks in the brain (e.g., Fleming & Dolan, 2012). For instance, patients with anterior prefrontal lesions show specific impairments in their metacognitive monitoring ability (Rounis et al., 2010; Ryals et al., 2016), while transcranial theta-burst stimulation has been shown to affect metacognitive monitoring without altering type-1 performance (Pannu & Kaszniak, 2005). At a neuroanatomical level, Fleming et al. have shown in an adult population that individual differences in metacognitive skill in the perception domain, are associated with interindividual variability in the volume of gray matter in the prefrontal cortex and also in the microstructure of white matter involving the same region (Fleming et al., 2010).

Developmental studies tackling these questions are scarce. Fandakova et al. were among the first ones to show that the prefrontal cortex has an important role also in children's metacognitive monitoring ability (7-12 years old), and linked changes of individuals' cortical structure during these years with their meta-memory skills (Fandakova et al., 2017). Next, in a large scale longitudinal study, Wendelken et al. (2017) suggested that the degree of the structural connectivity in the white matter between frontoparietal cortices during childhood (6 to 11 years old) can be predictive of cognitive functioning and reasoning skills, as well as of the degree of functional connectivity between these areas (Wendelken et al., 2017). Following these studies, Filevich et al. investigated whether these particular neural networks support the development of metacognition in early childhood, during the preschool to primary school transition. Findings of this study suggested that the age of 5 to 6 is a critical age window in the development of explicit forms of metacognitive monitoring. Using tasks in which children had to recognize and report their knowledge certainty in the task, they showed that children's ability to correctly report that they did not know is associated with key changes in cortical thickness in the medial orbitofrontal cortex (Filevich et al., 2020).

The first year of schooling (ages from 5 to 7 years old) has been suggested to bring remarkable changes in children's cognitive abilities and specifically their ability to control their behaviour. For instance, Brod et al. (2018) suggested that, during this year, students show great improvements in tasks requiring executive control functions, which are also linked to activity changes in parietal cortex regions associated with attention control. Moreover, they used fMRI to show that these children display increased activation of brain areas related to sustained attention during the first year of schooling and they attributed these changes to the structured environment of formal education (Brod et al., 2017).

The highlighted cognitive and neural changes occurring during the first years of primary school, together with the initiation of formal instruction to reading in the same age window, motivated us to study the role of metacognition during these early stages of reading acquisition. In the next section, we will discuss an outstanding question in the literature of metacognition regarding whether this involves a general process of monitoring or whether it is domain-specific, which would be a crucial issue to consider when designing educational interventions to train metacognitive skills in primary school.

1.1.5. Domain-specific/general processes underlying metacognitive monitoring

The question of whether metacognitive monitoring is supported by domain-general or specific mechanisms remains highly debated in the literature of metacognition. A domain-general model predicts that an individual with poor/good metacognitive ability in one domain (e.g., spelling performance), will have poor/good metacognitive skills in a different unrelated domain (e.g., recognizing emotions), hence supporting the view that a single metacognitive system monitors performance across different domains.

Several behavioural studies in adults support this model by showing that metacognitive performance is correlated across unrelated cognitive tasks (e.g., visual and auditory perception, and memory, see: Ais et al., 2016; Mazancieux et al., 2020, Preprint; McCurdy et al., 2013; Schraw et al., 1995). However, there are also contradictory results showing no correlation across perceptual and memory domains, or across tasks with distinct stimulus features (Samaha & Postle, 2017). Recent neuroimaging studies using perceptual and memory tasks in adults, suggest the presence of both domain-general neural structures in the frontal and posterior midline predicting confidence and accuracy in the task, and also domain-specific resources in the anterior PFC, responsible for assigning lower-level feelings of confidence to higher-order domain-related contextual cues (Morales et al., 2018; Rouault et al., 2018).

As discussed in the literature, potential discrepancies between behavioural studies could be due to the fact that (i) studies supporting domain-specificity may not have enough statistical power to detect correlations between domains and (ii) the inconsistency in the use of metacognitive indexes, with certain measures not controlling for the effect of metacognitive confidence bias or the confounding effect of type-1 performance (e.g., gamma and phi correlations, AUROC2; see Fleming & Lau, 2014) on metacognitive measures. This may lead to the detection of correlations of metacognitive sensitivity between domains, which are driven by type-1 task performance (Rouault et al., 2018).

Developmental studies assessing this issue are still very limited. Recent studies suggest that during middle childhood there is a gradual shift from domain-specific to domain-general resources supporting metacognition (Geurten et al., 2018; Lyons & Ghetti, 2010; Vo et al., 2014). However, it still remains unclear in which age this shift is occurring. For instance, Geurten et al (2018) suggested that this change is happening after the age of 10 (Geurten et al., 2018; Lyons & Ghetti, 2010; Vo et al., 2014), while Bellon et al. (2019) showed that already from the age of 8 metacognitive ability was correlated across domains (Bellon et al., 2019; Geurten et al., 2018). However, the results of these studies may suffer from the same factors of ambiguity presented in adult literature (small power to support non-significant correlations, metacognitive indexes that are confounded by the level of type-1 task performance, different task stimuli structure).

The current study will tackle the above issues by using the index of metacognitive efficiency, which mitigates confounds due to type-1 performance differences and confidence biases, hence allowing for cross-task comparisons (Fleming, 2017). Investigating how this system works during the first year of reading acquisition will help researchers and educators understand whether metacognitive ability can be boosted holistically, i.e., across domains, or whether the development of metacognitive strategies related to reading acquisition should be assisted separately.

1.2. Reading development and metacognition

1.2.1. Theories of reading development and models of reading

Reading acquisition requires the employment of a complex set of cognitive and linguistic skills (i.e., low-level visual processing, phonological processing, higher-level linguistic processes to access the semantics of printed words), which develop from an early age, before and while the individual learns to read (Georgiou et al., 2012; González-Valenzuela et al., 2016). There is a wealth of literature focusing on the extent to which these skills influence reading skillfulness, the developmental stages in which an individual acquires them, and the essential components and processes needed for one to transform from a novice to a skilled reader, who decodes and comprehends efficiently a printed text.

In the present study, we focus on the process of visual word identification, which is a crucial skill a reader needs to develop in the initial stages of reading acquisition. A novice reader requires a large amount of cognitive resources in order to identify a single word, which requires mapping graphemes to phonemes or larger grain sizes onto their corresponding phonological representations (Ziegler & Goswami, 2005). Mastering those mappings and repeatedly being exposed to words that become familiar to the reader, allows for the acquisition of a large repository of lexical orthographic representations, facilitating automaticity in visual word recognition in more skilled readers (Share & Shalev, 2004; Share, 1999). Automatic word recognition is a rapid process, which permits the reader to free up cognitive resources for higher-level linguistic processes, like the comprehension of a text (Ehri, 2005; Verhoeven & Perfetti, 2017).

For the purposes of the present study, we will first present the most prominent theories of reading development in order to understand the developmental stages in which students begin to master visual word identification, and second, the most commonly used computational models, that serve for breaking up visual word recognition into smaller components, in order to better understand the underlying cognitive operations supporting this process in novice and skilled readers (Castles et al., 2018).

1.2.1.1. Theories of reading development

The theories of reading development provide a theoretical framework for the different abilities a novice reader develops in each stage of reading acquisition in the process of learning how to read words by sight, automatically and rapidly. Here, we briefly describe these phases of reading development based on two commonly used reading models (see details in Ehri, 1995; Frith, 1985), which refer to typically developing readers.

During the *pre-reading phase*, a child has not yet received any formal instruction in reading, however, they do start to perceive stimuli of written text existing in their environment and begin to identify and memorize a few words as visual cues (e.g., the word STOP in the traffic sign). However, during this stage, children do not map letters into sounds and have not yet developed any orthographic skills. This stage has been defined as the *pre-alphabetic phase* of reading acquisition (Ehri, 1995, 2005) or the *logographic stage* (Frith, 1985).

Next, in the *early reading or early decoding phase,* when formal reading instruction starts to take place, Ehri (1995) has suggested that children begin to make some associations between letters or combinations of letters and their corresponding sounds and start being aware of the "orthographic principle", i.e., the understanding that the visual objects that are letters correspond to *linguistic* units. During this phase, which has been called the "*partial alphabetic phase*", children use these associations (usually recognizing the first and last letter of a word) in order to decode or "guess" a written word. However, they can still get easily confused when pronouncing a word (e.g., for the written word *ball,* pronouncing /bal/ instead of /bol/) or when writing a spoken word (e.g., writing bol instead of ball, see: Ehri, 1995).

As individuals learn to map graphemes to phonemes, the addition of words to their orthographic repository increases. When these mappings are mastered, with repeated exposure, students start to visually identify whole word forms (sight word reading). This stage has been defined as the "*full alphabetic*" stage by Ehri (1995) or the "*alphabetic*" stage by Frith (1985). Frith's alphabetic stage includes all processes occurring in the partial and full alphabetic stages of Ehri's theory (see Figure 2, Ehri, 1995; Frith, 1985).

During the final stage of reading development (*orthographic stage*; Frith, 1985 or *consolidated alphabetic stage*; Ehri, 1995), students encounter familiar written words with increasing frequency and start decoding whole word forms using larger grain sizes of orthographic representations. Moreover, it has been suggested that during this stage students show an increasing ability to extract orthographic patterns/regularities. Students make use of these statistical regularities in order to process unfamiliar words to them (e.g., pronouncing the words *ball, call, mall*, see: Ehri, 1995; Frith, 1985).

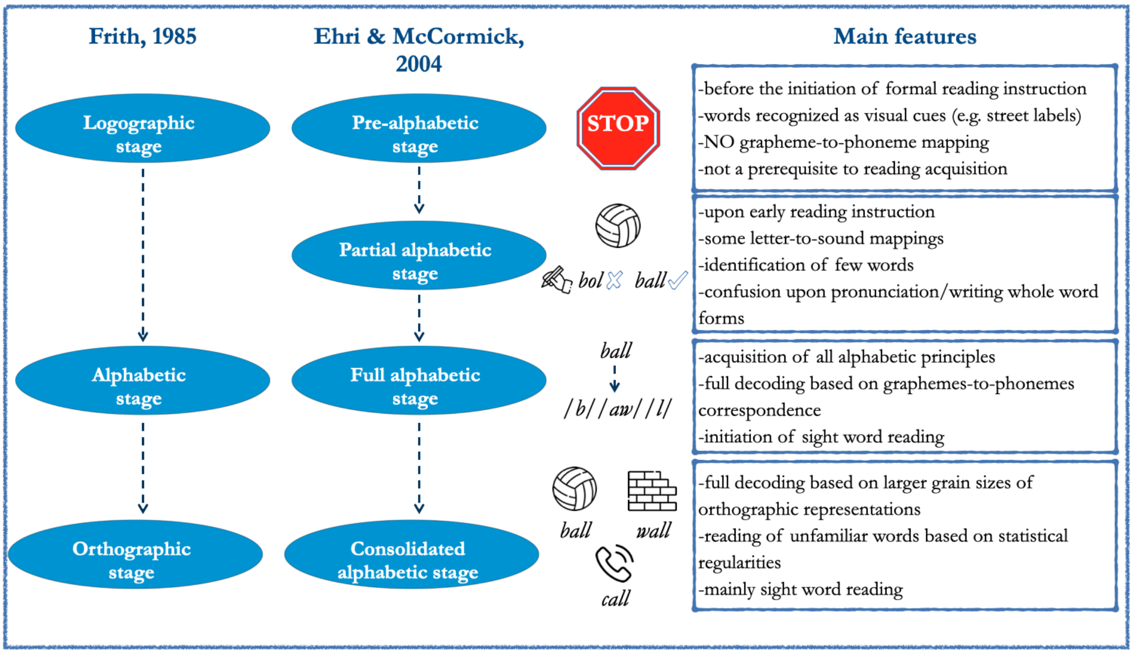


Figure 2 Phases of reading development and their main features (adapted from Ehri, 1995; Frith, 1985).

An alternative theory of reading development, called the “self-teaching” hypothesis, was later provided by Share (1995,1999), suggesting that sight word reading develops continuously from the very early stages of reading acquisition and that it is self-taught. Specifically, Share (1995, 1999) suggested that students receive explicit instructions on how to map graphemes onto phonemes, but that orthographic representations of whole word forms are established in the orthographic lexicon depending on the exposure of a reader to a word and its phonological recoding. Knowledge obtained from the phonological recoding of familiar words can then be used by the reader to self-teach unfamiliar orthographic word representations (Share, 1995; Share, 1999).

1.2.1.2. Computational models of visual word recognition

In the field of Cognitive Psychology, transforming theoretical frameworks of reading into computational models has served to break the process of reading into different components that can be more easily investigated and tested in humans. *Visual word recognition* is an important component of reading, and understanding the underlying processes which allow a fluent reader to efficiently recognize a word can provide valuable information regarding how readers then comprehend whole phrases or texts. Models of visual word recognition in fluent readers have taken two main approaches, proposing either a distinct mechanism for familiar and unfamiliar word reading (dual route models), or a connectionist network used to read both familiar and unfamiliar words.

The Dual Route Cascade (DRC) model is the most prominent example of a dual route model (see Figure 3, Coltheart et al., 2001; Coltheart & Rastle, 1994). This model suggests that familiar words are processed through a faster, lexical route, while unfamiliar words are processed by a non- or sub- lexical route, which is a distinct but parallel to the lexical route process. When a reader encounters a familiar word in a printed text, this sequence of letters directly activates the representation of this word in their orthographic and subsequently phonological lexicon, allowing them to read this word aloud. This route is in line with the process of sight word reading, which develops after repeated exposure to a word, and it is more predominant in fluent reading stages. On the other hand, when a reader encounters an unfamiliar word, then this needs to first be serially decoded using the sub-lexical route, by applying grapheme-to-phoneme conversion rules.

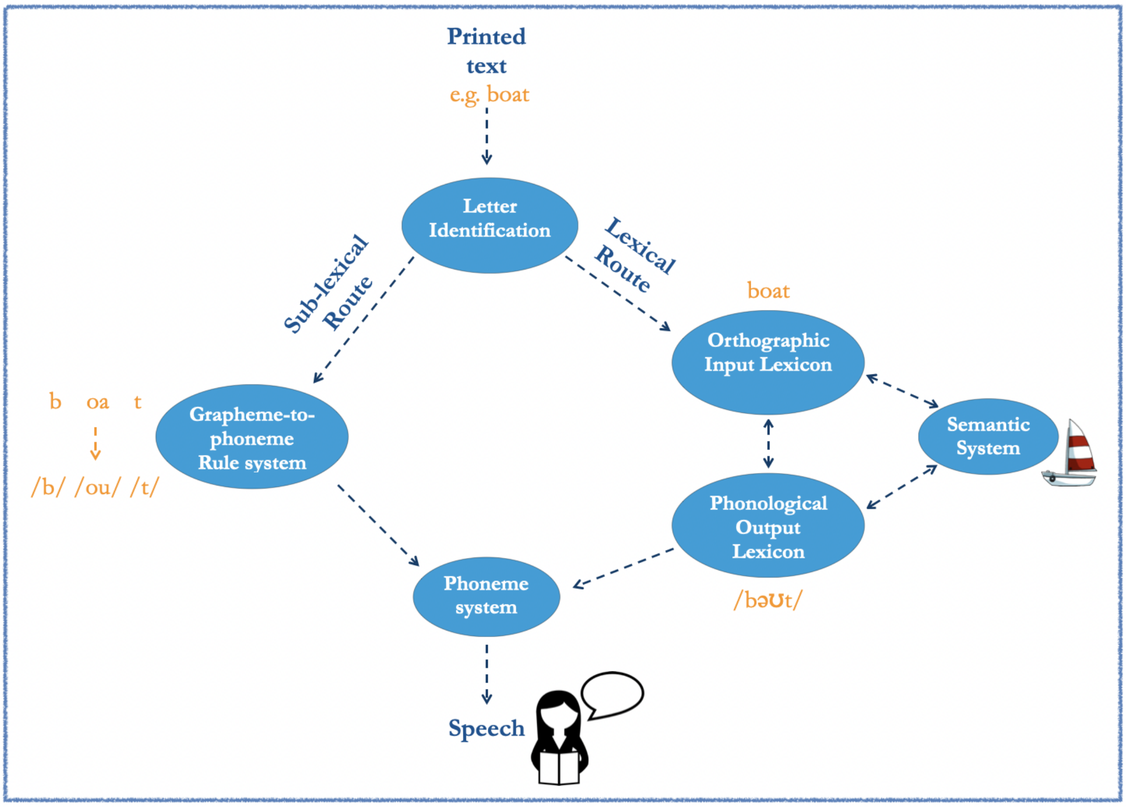


Figure 3 The dual route cascade (DRC) model of visual word recognition (adapted from Coltheart, 2006).

When decoding an unfamiliar word has been successfully accomplished, readers can use the lexical route in order to create an individual entry of this word in their orthographic lexicon, and access its meaning. Early readers rely more on the sub-lexical route (e.g., during the partial and alphabetic stages, Ehri, 2005), as they depend on the decoding of words through grapheme-to-phoneme mappings (Coltheart et al., 2001; Coltheart & Rastle, 1994).

A disadvantage of the DRC model is that it does not explain how a reader can process exception words, which means words that do not follow the grapheme-to-phoneme rules of a certain language (e.g., the word '*sugar*' starts with an 's', but is pronounced as an /sh/). Moreover, the processing of non-words (e.g., '*bave*') by the DRC models has been previously challenged. Non-words can be processed and read-aloud using the sub-lexical route, but it is unclear whether they can enter the lexical route at all, or whether they do so by activating entries in the orthographic lexicon with similar properties (Glushko, 1979).

The triangle or parallel distributed processing (PDP) model was developed as a main opponent of the DRC model, aiming to tackle these issues (Harm & Seidenberg, 2004; Plaut et al., 1996). The PDP involves a connectionist model which uses the same mechanism to process familiar and unfamiliar words (see Figure 4, for a review see: Rayner & Reichle, 2010). Based on the PDP model, reading involves the interaction between three key systems (i) the orthographic lexicon (ii) the phonological lexicon, and (iii) a semantic processing system.

A lexical entry entering the model activates in parallel all three units of the system, and all hidden layers inside each system (distributed representation). For instance, the representation of the word *ball* will activate all existing representations in the lexicon including the larger grain *-all* (e.g., *hall, mall, tall).* Hence, information about an entry is represented in the model like in a neural network; once this entry is activated, all the related parts of the network activate as well (see Figure 4).

In contrast to the DRC model, the PDP model does not depend on grapheme-to-phoneme conversions to identify a word but processes a word as a whole. With increasing exposure to lexical entries, the network of the different units becomes stronger and activates faster. Moreover, by exposure to different lexical entries, the model can start detecting statistical regularities at a sub-lexical level, both in orthography and in phonology units (Harm & Seidenberg, 1999).

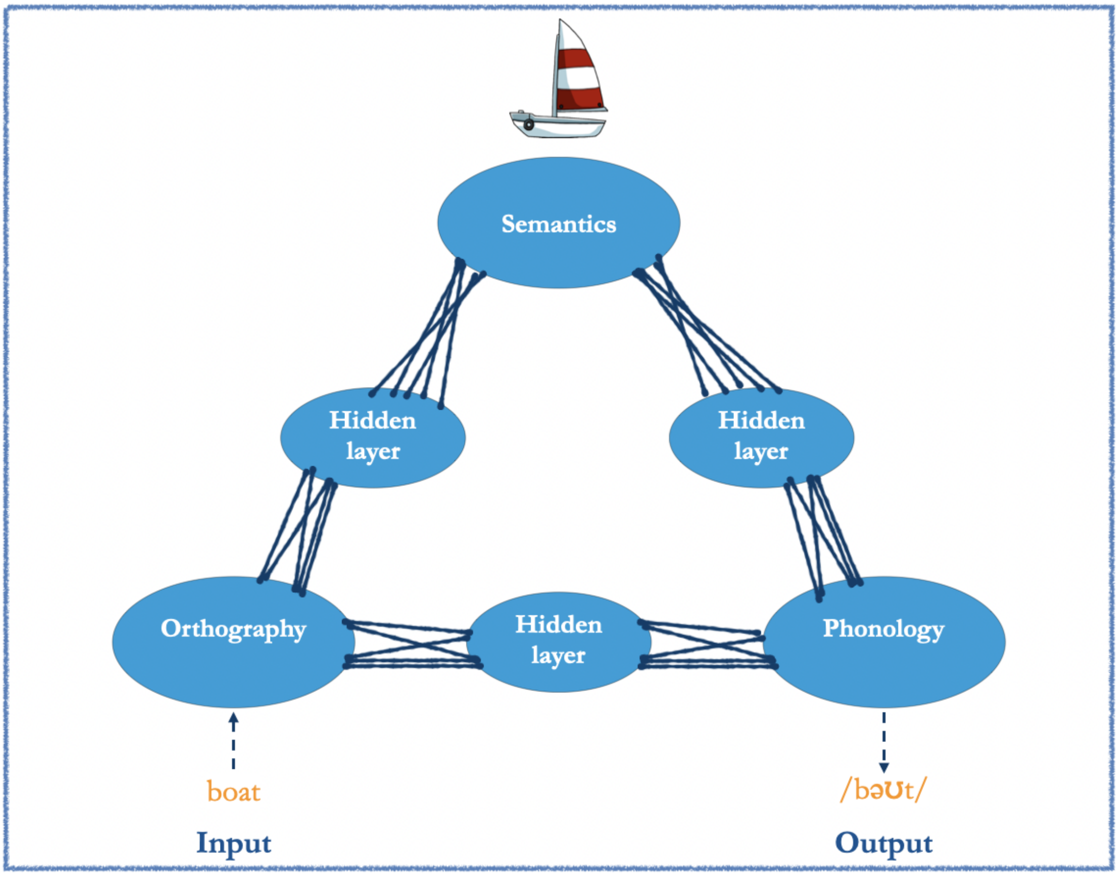


Figure 4 The triangle or parallel distributed processing (PDP) model of visual word recognition (adapted from Coltheart et al., 2010).

These statistical regularities can be used for the processing of unfamiliar words and non-words through the same dynamic system. Overall, although the DRC model has been shown to be more efficient in modelling human performance in skilled readers than the PDP model, it does not explain how reading is learned in the first place. The PDP model offers an account of how reading may be learned, but it has not yet been proven that children learn in this way (see Coltheart, 2006, for a contrast of the two models). However, both models can be useful tools when studying the relationship between their component processes during visual word recognition.

Following the above-mentioned influential models, hybrid reading models have been proposed in the reading literature attempting to combine the strengths of the dual-route and connectionist models and further explain the mechanism of skilled reading. For instance, the *multiple trace model (MTM) of polysyllabic word reading* (Ans et al., 1998), is a connectionist model, which suggests the existence of two reading procedures; *global reading*, which arises first, usually upon familiar word reading, and *analytic reading*, which is applied only when global reading fails. The MTM, unlike the previously proposed models, was the first one to integrate visual attentional processing of printed text as a crucial process in skilled reading. The model predicts that upon the presence of an orthographic input (orthographic layer 1, see Figure 5), the reader can simultaneously process a specific amount of visual orthographic information, depending on their familiarity with the input and/or their visual attentional capacities. This amount of visual information extracted is referred to as an individual's "visual attentional window". Precisely, Ans et al. (1998) suggested that participants' visual attentional window assists *global reading* by expanding over multiple elements and allowing their parallel processing, or narrows down allowing the individual to process a lexical entry in several steps by doing grapheme-to-phoneme mappings, during *analytic -e.g., syllabic- reading* (see Figure 5, Ans et al., 1998; Bosse et al., 2007).

Based on the MTM, global reading is benefitted when an individual has a larger visual attention window. Moreover, when global reading is successful, it allows readers to retrieve word traces from their episodic memory and output the whole word orthographic representation in orthographic layer 2 (orthographic layer 1 input = orthographic layer 2 output, see Figure 5).

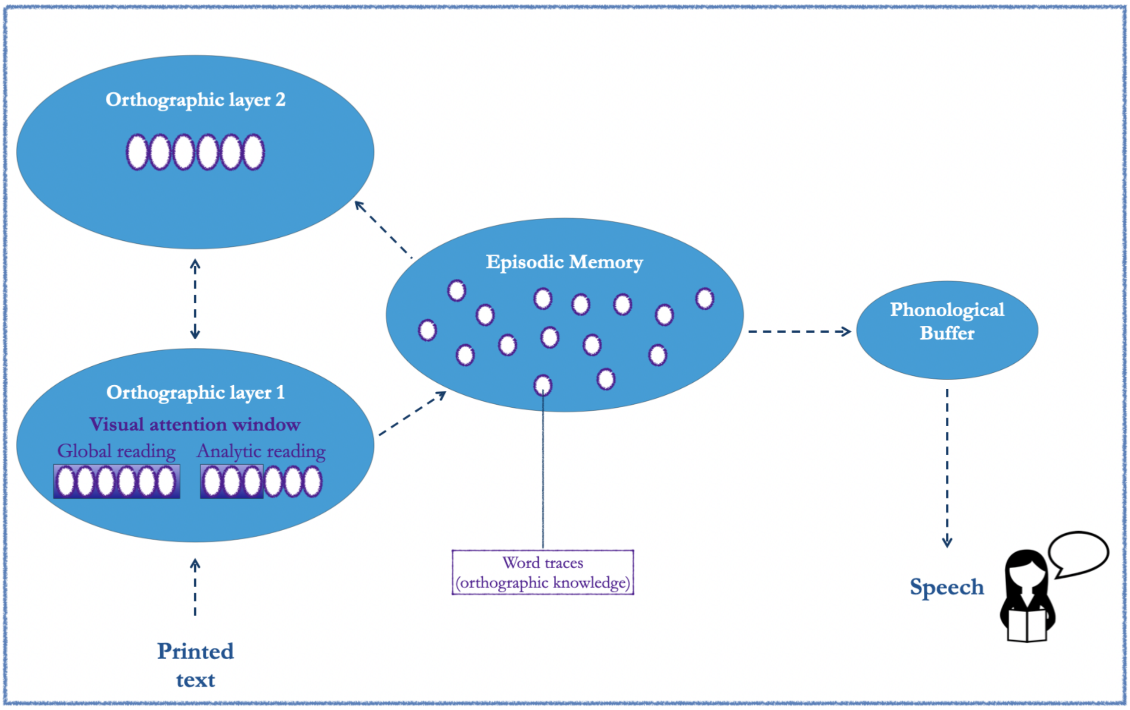


Figure 5 The multiple trace model (MTM) of polysyllabic word reading (adapted from Ans et al., 1998).

On the contrary, when the word is not familiar, the global mode fails (i.e., orthographic layer 1 input ≠ orthographic layer 2 output, see Figure 5), and the attentional window narrows down to perform analytic reading that allows the reader to input in the orthographic layer 1 smaller grains. These smaller orthographic (sublexical) units get activated in the episodic memory and decoded through a visual attentional sequential scanning. When the sequential processing of sublexical units composing a reading item is completed and is simultaneously available with the whole item phonological representation, a new word-trace can be created in the episodic memory and outputted in the phonological buffer in order to be read aloud. Interestingly, this analytic mode will also be at play when an individual does not have a large enough visual attention window to process a whole word form (Ans et al., 1998; Bosse et al., 2007). Overall, MTM highlights the importance of individuals' visual attention in reading, which can influence and contribute both in lexical and sublexical reading, and the strengthening of orthographic representations in readers' orthographic lexicon.

In the following section, we will discuss the role of orthographic lexical processing, an important component of all models mentioned above, during reading development.

1.2.2. The role of orthographic lexical processing and its prerequisites during reading development

This study specifically focuses on understanding the role of metacognition in the development of individuals' orthographic lexicon, which refers to the orthographic knowledge and processing of whole word forms and, as mentioned in the previous sections, it is essential for individuals to develop fast and fluent reading (Ehri, 2014; Frith, 1985).

Previous literature in reading acquisition was initially focused on the importance of the development of readers' phonological skills upon the establishment of their reading system (Anthony & Francis, 2005; Hudson et al., 2008). However, recent lines of research are now emphasizing the crucial role of visual processing of the orthographic form of words in influencing lexical access during visual word recognition (Bosse et al., 2007; Bosse & Valdois, 2009; Kwon et al., 2007). A sequence of letters in a printed text needs to be accurately identified first visually, in order to then be transformed into sounds, and give access to the meaning of words, if this exists in the semantic system of an individual.

As per the MTM model mentioned in Chapter Section 1.2.1, individuals' visual attention (VA) window has been found to impact reading by allowing the distribution of attentional resources in processing series of letters simultaneously, in order for one to be able to identify whole orthographic forms through the global reading mode (Ans et al., 1998; Bundesen, 1990). In this context, the visual attention window has been operationalized under the cognitive concept of the VA span, which refers to the number of distinct visual elements (e.g., letters, symbols, numbers) that can be identified in parallel under one eye fixation. This simultaneous processing of strings of letters has been shown to facilitate the formation of lexical orthographic traces in memory for future access (Bosse et al., 2007; Bosse & Valdois, 2009).

VA span is a pre-orthographic processing skill, which has been found to predict not only single word reading, independently of individuals' phonological awareness and decoding skills and their working memory ability (Bosse et al., 2007; Bosse & Valdois, 2009; Lallier et al., 2013), but also reading fluency (Besner et al., 2016; Valdois et al., 2019), and reading comprehension (Chen et al., 2016). Moreover, deficits in the VA span have been linked with difficulties in long-word or pseudoword readings in dyslexic and developing readers (van den Boer et al., 2013). Finally, VA span has been associated with students' reading aloud speed of any reading item, with irregular word reading accuracy (Bosse & Valdois, 2009; Germano et al., 2014), and with single word spelling abilities during the first years of primary school, both in shallow and deep orthographies (Ginestet et al., 2019; Niolaki et al., 2013).

Once an individual can visually process a printed word, they can subsequently access their orthographic lexicon in order to retrieve whole word orthographic representations from their memory (Chetail, 2017; Ginestet et al., 2019). The efficiency of finding an orthographic representation of a printed word in one's orthographic lexicon, and therefore, the quality of this representation (lexical orthographic knowledge), has been studied by measuring individuals' performance in a lexical decision task. In this, one has to categorize briefly presented letter sequences into words or non-words. Lexical orthographic knowledge has been suggested to provide the link between the employment of visual attention resources to process a printed word, and reader's access to the phonology of this word, which is crucial for sight word reading (Stanovich, 1993). However, there is still little research providing evidence of the direct link between the VA span and participants' performance in the lexical decision task. Holmes and Dawnson (2014) reported that the width of adult readers' VA span has little contribution only in the classification of regularly, but not irregularly, spelled words composed of more than 5 letters, and of short (4 or 5 letters in length), but not long pseudowords (more than 5 letters in length), while Ginestet et al. (2019) later suggested that participants' VA span can modulate length effects for words in a lexical decision task (Holmes & Dawson, 2014;Ginestet et al., 2019).

Individuals' performance in the lexical decision task has been found to significantly predict word identification skills and also to explain substantial variance in the reading aloud accuracy of words, but not in reading comprehension (Katz et al., 2012).

Overall, the abovementioned skills (VA span, orthographic lexical knowledge) have been considered essential in the process of visual word recognition and for allowing the development of automatic or sight word reading. An increase of an individuals' automaticity in visual word recognition has been suggested to free up cognitive resources, which can in turn be employed in reading comprehension (Verhoeven et al., 2019). For this reason, automaticity in word recognition has been considered to be an important predictor of academic achievement in reading (Cunningham & Stanovich, 1997).

Understanding how different cognitive and linguistic factors can contribute in the development of visual word recognition can be critical, both for understanding better the underlying reading processes and updating the existing reading models and theories, but also towards the faster assessment and subsequent implementation of interventions in individuals with reading difficulties, from the early stages of reading acquisition. In the present study, we focus on investigating the contribution of metacognitive monitoring, a higher-order thinking skill, on tasks assessing prerequisites of orthographic lexical processing and visual word recognition (i.e., the VA span, efficiency of individuals' lexical orthographic knowledge, sight word reading).

In the following section, we will review the up-to-date literature related to metacognition and reading performance.

1.2.3. Previous studies in metacognition and reading development

During the last decades, educational studies increasingly highlight the importance of self-regulated learning in academic achievement, involving the active engagement of students in their own learning processes. The process of metacognitive monitoring and subsequent regulation of learning, has been considered an important prerequisite of students' self-regulated learning, together with the motivation and cognitive abilities of the students (Efklides, 2008; Zimmerman & Moylan, 2009). For instance, if a student was studying for an exam, by being able to efficiently monitor their already acquired and well-established knowledge, they would know how much studying is enough, or in which items they would need to focus more on re-studying.

Given that most of the information a student has access to is in the form of written text, metacognitive monitoring in relation to reading has arisen a special interest in researchers. However, up-to-date studies are mostly limited to studying metacognitive monitoring in reading in relation to comprehension, which has been defined as "metacomprehension" (Garner, 1987; Jacobs & Paris, 1987; Roebers et al., 2009). Moreover, most of the studies on this field entail the limitation that they use metacognitive indexes that either derive from self-report questionnaires, or do not use robust metrics of metacognition, which are not confounded by students' confidence bias and type-1 performance. Below, we review the existing literature and subsequently support the added value of our study within this framework.

As mentioned in Chapter Section 1.2.1., reading comprehension is a notably complex skill, involving several linguistic competencies (e.g., orthographic processing, phonological recoding) and cognitive skills (e.g., visual attention, working memory, critical thinking, ability to make inferences) (for a review see Verhoeven et al., 2011). In the context of reading comprehension, metacognitive monitoring allows an individual to judge how well they have understood a whole text, detect the passages to focus on in order to increase their comprehension, and subsequently activate their metacognitive control system in order to improve their learning.

Monitoring of reading comprehension was first studied using the error detection paradigm. In this, students were asked to detect specific inconsistencies in a text, e.g., grammatical or spelling errors and contradictory sentences (Winograd & Johnston, 1982). Using this paradigm, several studies reported that primary school students with poor reading comprehension abilities were struggling more to detect inconsistencies in a text, when compared to students with strong comprehension abilities (August et al., 1984; Oakhill et al., 2005; Otero & Kintsch, 1992). However, even though error monitoring and postdictive judgments (JOLs, confidence judgments) may imply common mechanisms, the latter have been suggested to provide a finer-tuning of human behaviour and optimization of learning (Yeung & Summerfield, 2012).

Few studies have yet examined the role of metacognitive monitoring using postdictive judgments during reading development. Moreover, most of them were based on the collection of JOLs, involving asking a participant to study a text and subsequently predict their future accuracy in upcoming questions on the text (e.g., 'How many of the questions presented do you think you will answer correctly?'), while few of them used confidence judgments asking participants to judge their certainty on responding correctly on a comprehension test.

Schneider et al. (2000) using this paradigm, suggested that children of different age groups (kindergartner, Grade 2, Grade 4) were more efficient in providing JOLs if there was a delay between the reading of the text and the collection of JOLs responses (Schneider et al., 2000).

Roebers et al. (2009) using a paper and pencil test concerning comprehension questions, and subsequent rating of confidence, on a previously watched film related to sugar production, reported that 9 and 11 years old students were found to be equally good at their metacognitive monitoring skills regarding their comprehension of the content of their movie. However, the 11 years old children were found to be better at their metacognitive control skills, as they were more accurate in withdrawing incorrect answers when this was given as an option (Roebers et al., 2009). Metacomprehension ability of children of different age groups (Grade 4, 6, 7) was estimated in a following study, by asking participants to read 5 short texts (350 words) and provide JOLs in tests including questions for each of the 5 texts separately. Subsequently, students' metacognitive control ability was assessed by asking participants to choose whether they will re-study the text or not. Moreover, half of the participants were asked to create keywords after reading each text, while the other half was not. Results of this study showed that students' metacognitive monitoring and subsequent control in Grade 6 and 7, but not in Grade 4, was better in the keyword condition (de Bruin et al., 2011). Here, it is important to mention, that in both of the previously mentioned studies, Goodman-Kruskal gamma correlations were used to calculate students' monitoring accuracy, a metacognitive index which has been criticized for not accounting for participants' tendency to be over or under confident (Masson & Rotello, 2009). In the present study, in order to overcome this issue, we will use a hierarchical Bayesian SDT model in order to provide free-of-biases metacognitive monitoring estimates.

Overall, previous studies in typically developing populations pinpoint the role of metacognition in reading comprehension, especially after middle childhood, which, based on the reading models mentioned in Chapter Section 1.2.1., correspond to ages when children usually develop fluent reading. The present study is, to our knowledge, the first to examine the role of metacognition in visual word recognition and orthographic processing, which are skills that develop early on during reading development, and even though they have been found to predict reading comprehension, they involve different cognitive processes. Understanding how and if metacognitive monitoring is involved in the development of these skills can provide useful insights on how to support early readers in an educational context. In the following section, we will outline the general objectives of the present thesis, and the methodology employed to achieve them.

1.3. The present thesis

1.3.1. Objectives

The previous sections of the General Introduction of this thesis outlined the definition of metacognition, including up-to-date knowledge regarding its developmental trajectories and the contribution of metacognition to students' learning in different domains. Moreover, the classical models of reading were presented and the contribution of orthographic lexical processing in reading performance was discussed.

The present study will employ a longitudinal approach to assess students' metacognition in tasks related to orthographic lexical processing and in a non-reading related task, as well as students' standardized reading performance in the first years of primary school (Grade 1 to Grade 3). The goal is to provide new insights in the following topics:

1. The development of metacognition in tasks related to orthographic lexical processing (vs a non-linguistic task) during the first years of primary school.
2. The use of domain general or domain specific mechanisms supporting metacognition in tasks related to orthographic lexical processing, and a non-reading task, during early childhood.
3. The role of metacognition in regulating students' learning, namely, during the development of students' orthographic lexicon and reading ability.

These research aims will be addressed in three chapters (II, III & IV) with the following structure. In Chapter II, we will focus on the intercorrelations between participants' type-1 task performance in the linguistic and non-linguistic experimental tasks and participants' standardized reading performance, as well as the development of these variables across the first years of primary school. By doing so, we will set the basis for later exploring the role of metacognition on the same tasks. In Chapter III, we will address the first and the second topic mentioned above, by investigating potential changes in metacognitive ability across the time points of the study and testing cross-tasks associations in students' metacognition in the linguistic and non-linguistic experimental tasks. Finally, in Chapter IV, we will test for associations between students' metacognition and task performance in the tasks assessing orthographic lexical processing within and across the first years of primary school.

1.3.2. Methodological Considerations

The present chapter will provide a general overview of the longitudinal experimental design of this study, the background and characteristics of the participants who took part in the study, and a description of the battery of tasks used across the different chapters. This information is located in this separate chapter for reasons of clarity and in order to avoid repetition in the following chapters. Below, a schematic representation of the design of this longitudinal study follows (see Figure 6).

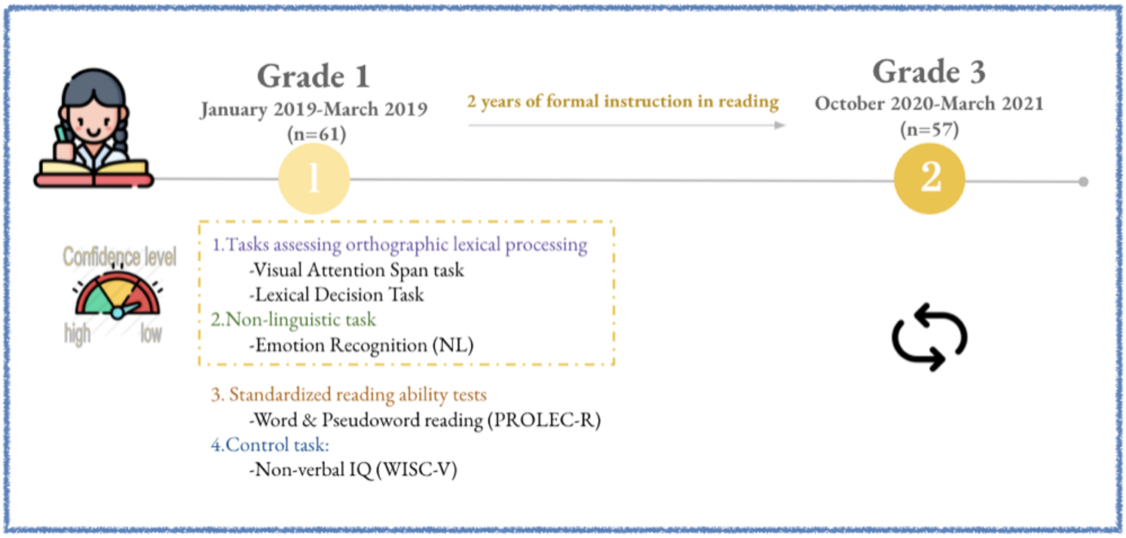


Figure 6 Schematic representation of the presented longitudinal study studying the role of metacognitive monitoring in early reading acquisition. The dashed yellow border indicates the experimental cognitive tasks for which both task performance (type-1) and metacognitive ability (type-2) were assessed.

1.3.2.1. Participants

Sixty-nine children aged between 6 and 7 years (mean age (±SD): 6.59 ± 0.29, 28 girls) attending Grade 1 (January 2019-March 2019) were initially recruited for this study. They were native Spanish speakers from an urban school in Vitoria, Spain. Nine participants were excluded from the analyses due to missing data in several tasks, due to inability to read, or due to low non-verbal IQ.

Participants were asked to participate voluntarily, and fully informed consent forms were obtained from the legal tutors of the minors, and the minors, prior to the study (see Appendix A for full information sheets and consent forms). The study was approved by the BCBL Ethics Board and the Bioethics Commission of the University of Barcelona (see Appendix B). Participants were retested 2 years later in Grade 3 (October 2020-March 2021, when the sample consisted of fifty-seven children (mean age (±SD): 8.48 ± 0.34, 25 girls). The reduction of the sample was due to outliers identified in the first time point of the study, that were not re-assessed in Grade 3, or participants' change of school. Participants' characteristics are presented in detail in Table 1.

Table 1 Participants' characteristics regarding their chronological age and non-verbal IQ in each time point of the longitudinal study. Mean Score (SD), Range, and Skewness measures are presented.

	Grade 1	Grade 3
Chronological age (years)		
Mean (SD)	6.59 (0.29)	8.48(0.34)
Range	6.14-7.14	7.92-9.08
Skewness	-0.070	0.075
Non-verbal IQ		
Mean (SD)	11.46 (3.24)	11.77 (2.64)
Range	6.00-19.00	6.00-18.00
Skewness	0.215	-0.002

1.3.2.2. Materials and Methods

Participants performed two linguistic tasks related to orthographic lexical processing and orthographic knowledge and one non-reading related task, in order to assess any domain-specific effects. The tasks were programmed in PsychoPy version 1.83.4 (Peirce & Macaskill, 2019) and were part of a larger task battery, which was administered in counterbalanced order during three sessions of 1 hour in participants' school. The schedule was organized in agreement with the teachers and director of the school. Each participant was tested individually in a soundproof room of the school to minimize noise. Participants sat in front of the computer screen at a distance of 70cm without head constraint. In addition to these tasks, participants were also administered two standardized reading tests (word and pseudoword reading) as well as a control task aimed at measuring their non-verbal reasoning abilities (non-verbal IQ).

1.3.2.3. Experimental tasks

1. Linguistic tasks related to orthographic lexical processing

Visual Attention (VA) Span Task

Visual stimuli were composed of 103 distinct 4-consonant strings (e.g., D P N L), created by the use of 13 consonants (B, D, F, G, H, K, L, M, N, P, R, S, T) (see Appendix C1 for detailed stimuli presentation). The present task was designed following well-established experimental protocols used in previous studies to measure VA span (Bosse et al., 2007; Bosse & Valdois, 2009). The criteria for the selection of the consonant strings were that no repetition of the consonants is permitted within the string, that strings do not contain grapheme clusters existing in Spanish language (e.g., ST, TR) and that they do not form word skeletons in Spanish (e.g., P L N T for "PLANETA"; meaning planet in Spanish). Visual stimuli of the present assessment were displayed on the computer screen using white upper-case Arial font with a black background. The 4-consonant strings occupied a space on the screen of min 5.3° and max 5.55° degrees of visual angle, with a 1.2 centre-to-centre distance between adjacent letters. This stimuli size was chosen in order to minimize lateral masking effects.

Each trial started with the onset of a central fixation cross at the center of the screen until the participant reported that they were ready to start the trial. The target 4-consonant string was displayed for 200 ms at the center of the screen. After the appearance of the string and following a blank screen of a 100 ms duration, a single target consonant was displayed in red font either slightly below or above the position that the 4-consonant string occupied (counterbalanced between trials). Participants were then asked to give a YES/NO response on whether the single consonant was part of the string or not by pressing keys on the keyboard labeled as ✓ or ✗. Subsequently, in each trial, participants were asked to rate their confidence on having given a correct response or not. Two options (1: I have doubts on whether my response was correct or not, 2: I'm sure my response was correct) were given to the participants that were explicitly explained in each trial. This was recorded through an external keyboard by the researcher (see a visual representation of a trial in Figure 7).

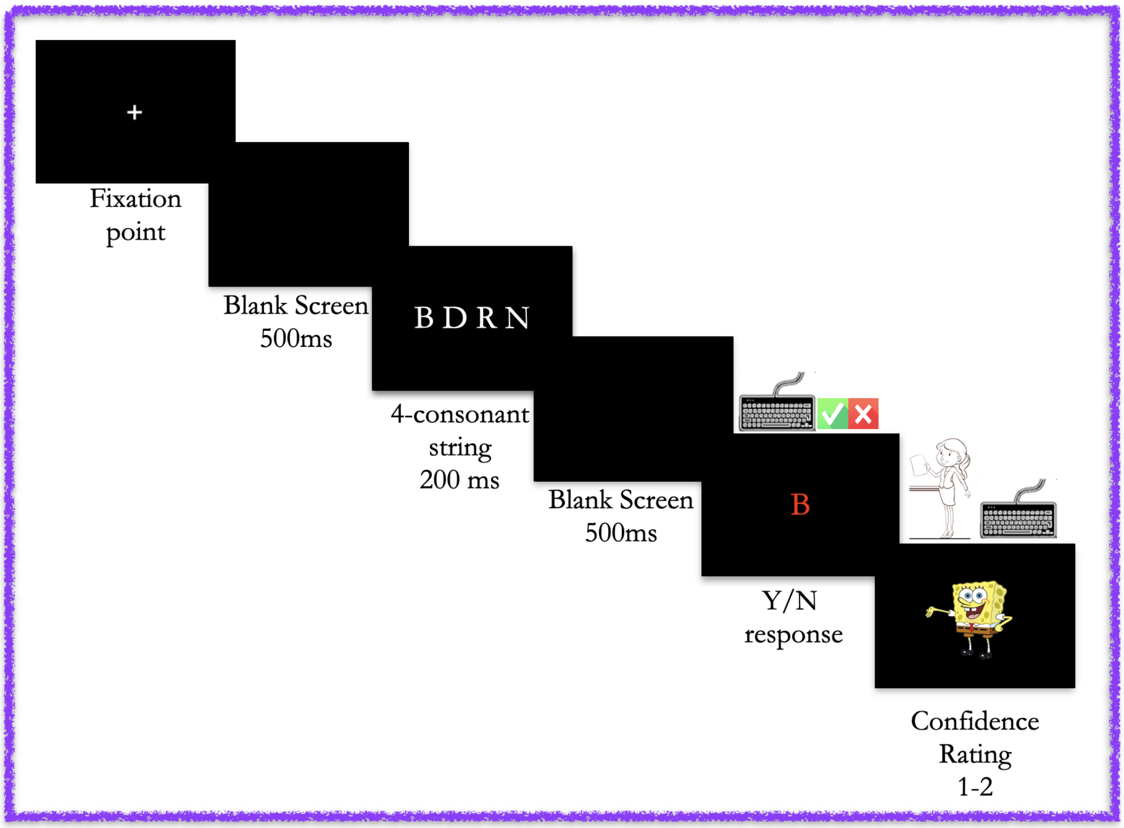


Figure 7 Behavioural task design of the Visual Attention Span task. Participants saw a briefly presented sequence composed of 4 consonants, followed by the presentation of a single target consonant in red. Participants had to decide whether the target consonant was part of the sequence (type-1 task) and rate their confidence upon this response (type-2 task).

Orthographic Lexical Decision Task

This task involved the visual presentation of 40 words and 40 pseudowords. Stimuli were selected from EsPaL (Duchon et al., 2013), a Spanish lexical database (see Appendix C2 for detailed stimuli presentation). The selected high-frequency words were 4-letters long. Orthographically legal pseudowords were created by changing one letter of words that were not those presented in the task, but with the same characteristics. The task was divided into two blocks of 40 items (counterbalanced between words and pseudowords) performed in separate sessions in order to ensure participants' attention to the task. Each trial started with the onset of a central fixation cross at the center of the screen until the participant reported that they were ready to start the trial. Following a blank screen of 500 ms, the target stimulus was presented. Target duration was calibrated for each participant prior to the experimental trials (see below). A backward mask (######) was then presented until participants' made their type-1 response, reporting whether the stimulus was a word or a non-word by pressing keys on the keyboard labeled as ✓ or ✗. Participants were then asked to rate their confidence on the response by using two options; 1: I have doubts on whether my response was correct or not, and 2: I'm sure my response was correct. This was recorded through an external keyboard by the researcher (see a visual representation of a trial in Figure 8).

In this task, a continuous staircase procedure was used to adjust stimulus presentation duration in order to avoid ceiling effects in type-1 accuracy and to equate participants' type-1 accuracy at around 70%. Such a procedure allows for more accurate estimates of type-2 metacognitive efficiency. In order to define the starting duration of each participant, a calibration phase resembling the characteristics of the main task was carried out first, in which the cumulative accuracy on the task was calculated after each trial. The stimulus presentation duration in which participants achieved approximately 70% accuracy on the task was used as a starting duration for the main task. During the experimental trials, the stimulus duration for the next trials was calibrated using a 2 down-1 up staircase, adapting to whether participants responded correctly or not. In the case of two sequential correct responses, the stimulus duration was decreasing by 1 frame, while in the case of one incorrect response, the stimulus duration was increasing by 1 frame.

Participants were assessed again in the same task 2 years later using a different set of stimuli (words and pseudowords) with the same characteristics.

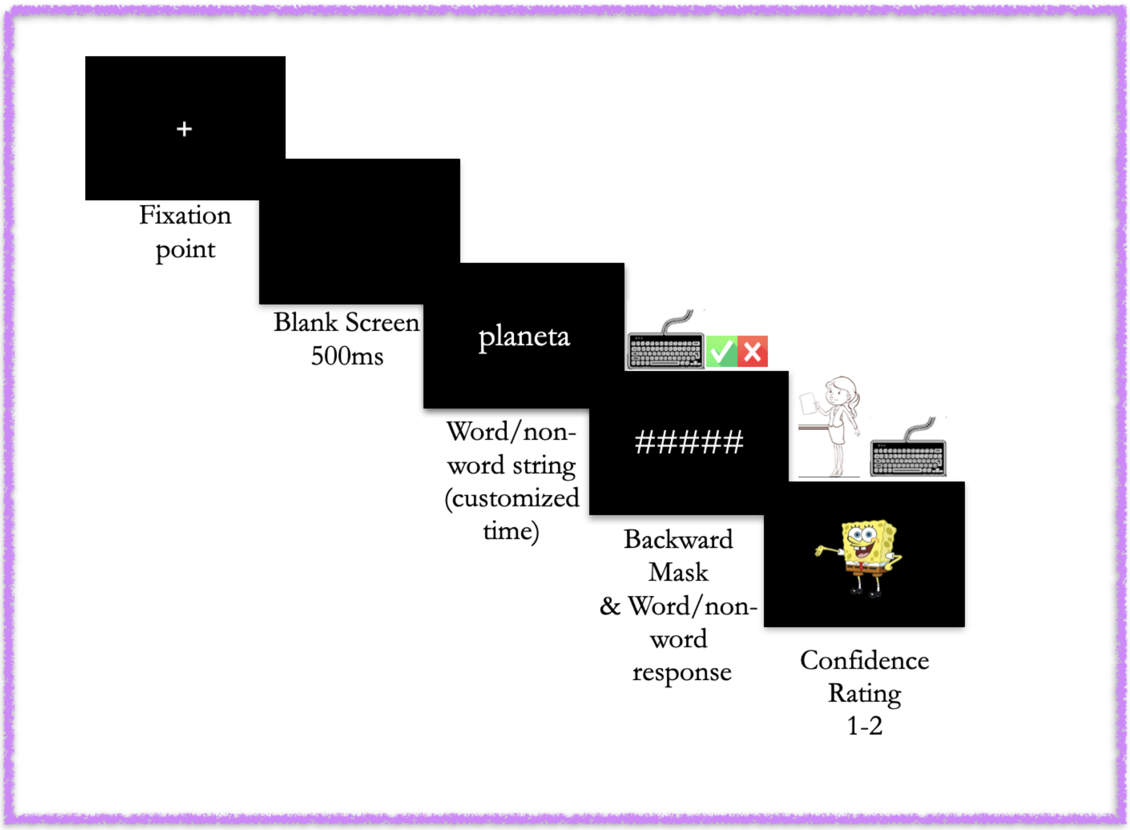


Figure 8 Behavioural task design of the Orthographic Lexical Decision task. Participants saw a briefly presented sequence of letters that composed words or pseudowords. Participants had to decide on the identity of the sequence (word vs pseudoword, type-1 task) and rate their confidence upon this response (type-2 task).

2. Non-linguistic task unrelated to reading skills: **Emotion Recognition Task**

An emotion recognition task was developed including the presentation of human face pictures for which participants were asked to make a decision on whether the face was happy or neutral. A total of 48 stimuli were presented, selected from the "Developmental Emotional Faces Stimulus Set" (DEFSS) (Meuwissen et al., 2017), which were balanced for mood, gender, and age (child or adult) of the face (see Appendix C3). Each trial started with a central fixation cross until the participant was ready to start the trial. The experimenter then initiated the trial. Following a blank screen of 500 ms, a forward mask (composed of different colours and cartoon robot pieces) appeared for 100 ms, and then the target stimulus appeared with a duration that was pre-calibrated prior to the experimental trials using a similar procedure to the orthographic lexical decision task above. After the offset of the stimuli, a backward mask was displayed on the screen and participants were asked to give a HAPPY/NEUTRAL response (by pressing keys on the keyboard labeled as 😊 or 😐) on whether the face they saw on the screen was depicting a happy or a neutral emotion. Subsequently, in each trial, participants were asked to rate their confidence in the same manner as outlined above (see a visual representation of a trial in Figure 9). Participants repeated the current task as described above 2 years later.

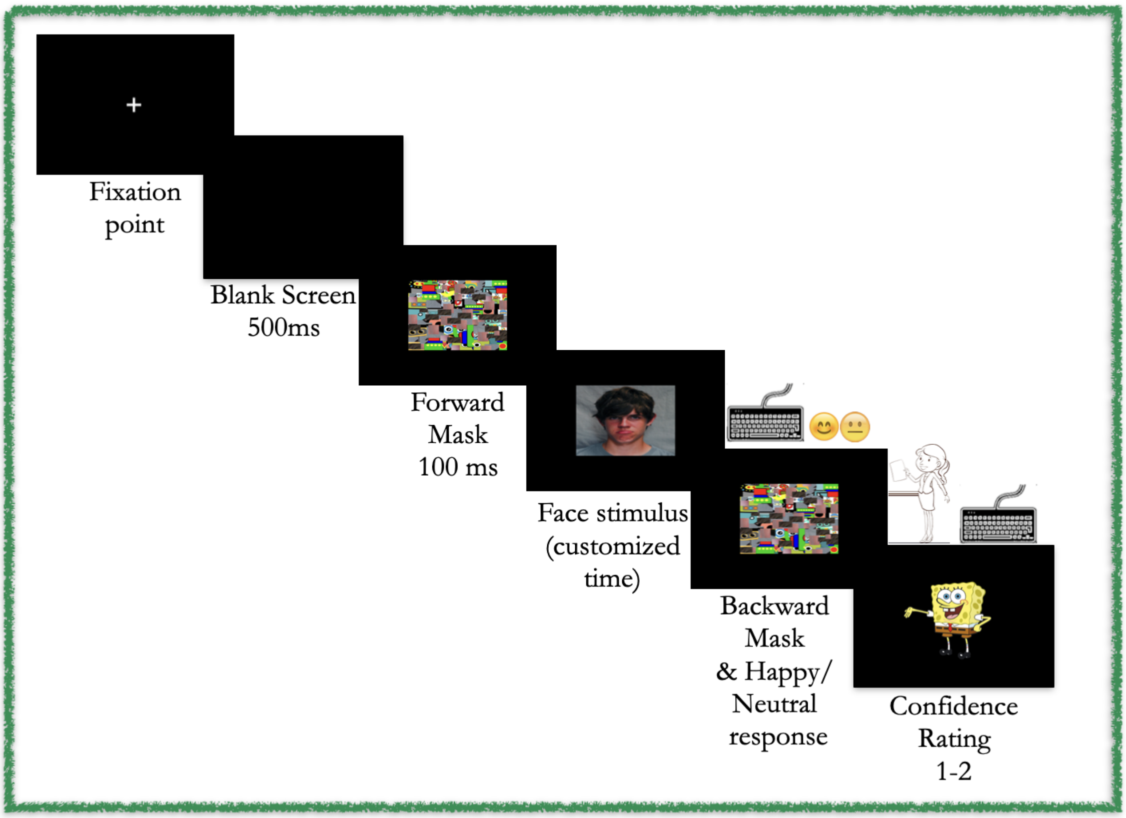


Figure 9 Behavioural task design of the Emotion Recognition Task. Participants saw a briefly presented face that expressed a happy or neutral emotion. Participants had to decide on the emotion of the face presented (happy vs neutral, type-1 task) and rate their confidence upon this response (type-2 task).

3. Standardized reading tests

In order to obtain a standardized score for tracking participants' reading skill, the single-item reading subtests of the PROLEC-R battery were used (see details in Cuetos, Rodrigues, Ruano, 1996). In this test, participants were given lists of words and pseudowords (40 items per list) and were asked to read them out loud (see Appendix C4 for the presented lists of words and pseudowords). The average time taken to read both lists in seconds, as well as the average percentage of correct responses was recorded and used as standardized reading measures in the present study.

4. Control task: Non-verbal IQ

Participants’ non-verbal IQ was assessed using the matrices subtests of WISC-V (Wechsler, 2014). The raw scores were first converted to scaled scores according to the age band each participant belonged to, following the tables of normative samples provided in the WISC-V manual. Participants with a score inferior to the 25th percentile on WISC-V matrices' scaled scores were excluded from the analysis. Scaled non-verbal IQ scores were used as a covariate in all analyses in order to rule out the possibility that given associations are driven by general factors of intelligence.

1.3.2.4. Preliminary data analysis

Prior to the main data analysis following in the experimental Chapters (II-IV), participants performing at chance level in a task were excluded from the analysis of the certain task (accuracy ≤ 0.5) at each time point of the longitudinal study. Subsequently, the interquartile range (IQR) criterion was used to screen for outliers. Values that were two interquartile ranges larger or smaller from the median were identified as outliers. Accuracy data were screened for outliers, and outliers were excluded separately in each experiment (see Table 2, Chapter Section 2.3.1., for No of participants included in each of the experimental tasks).

Our analyses focus on the estimation of metacognitive efficiency (meta-d’/d’) using an SDT framework. In SDT models, type-1 performance (d’ or d-prime) illustrates the ability of an observer to discriminate between two different states of the world (e.g., signal vs noise, word vs pseudoword, happy vs neutral). It is calculated as d’ = z (hits) - z (false alarms), where z(p), p $\in$ [0,1] is the inverse of the cumulative distribution function of the normal Gaussian distribution. “Hits” refer to the proportion of trials in which the subject detected ‘signal’ when the ‘signal’ was present, while “false alarms” refer to the proportion of trials in which the subject detected ‘signal’ and the ‘signal’ was absent. Type-2 SDT metacognitive performance (meta-d’) refers to the ability of the subject to discriminate between correct vs incorrect responses by means of the confidence ratings. Here in the type-2 analysis, “hits” correspond to the proportion of trials in which the subject responded with high confidence and the type-1 response was correct, while “false alarms” refer to the proportion of trials in which they responded with high confidence and the type-1 response was incorrect. Type-2 meta-d’ is estimated as the type-1 d’ value that would correspond to the observed confidence distributions in a metacognitive “ideal” subject (Maniscalco & Lau, 2012). Hence, type-1 d’ and type-2 meta-d' are in the same units, thus they can be comparable. In an ideal metacognitive observer: meta-d' = d'. If meta-d' < d', we can deduce that the subject is not using all the available stimulus information to inform their metacognitive system. In cases where meta-d' > d', subjects are supposed to further process the stimulus information fruitfully, after having given the type-1 response and before giving their metacognitive judgments.

Using meta-d' as an estimate of type-2 performance is free of confidence bias but it can be affected by the task difficulty. Calculating the ratio of meta-d'/d' (M-ratio) allows for an estimate of type-2 performance controlling for the subject's type-1 performance (reported in the results as "type-2 metacognitive efficiency"). This measure permits meaningful comparison across subjects or task domains.

In this study, HMeta-d toolbox (https://github.com/metacoglab/HMeta-d), a recently developed SDT hierarchical Bayesian framework (Fleming, 2017), was used to estimate metacognitive efficiency (meta-d'/d' or Mratio) in all tasks performed by the children. This framework allows for the estimation of type-2 performance both at a single-subject and a group level. Group level estimates handle subject-level uncertainty, meaning that a participant with a high level of uncertainty doesn't contribute equally to the estimation of group level parameters. These estimates are particularly useful for a direct estimation of covariance in metacognitive efficiency across tasks or time points. Also, HMeta-d Bayesian framework avoids edge correction, handling naturally possible zero cell counts in a certain confidence level. This was crucial in the current study, as in early childhood, several participants have a tendency to respond with a high confidence rating in a high proportion of trials, despite being instructed to use all the confidence ratings accordingly.

Normality tests were conducted in each variable of interest of the experimental tasks. The measure of Skewness was used to evaluate normality. Considering that many values were moderately to highly skewed (Skewness > 0.5 or < -0.5), we elected to use non-parametric tests for the frequentist testing, when this option was available or to use log-transformation to address skewed-data. A detailed description of the data analysis plan used in each chapter can be found in the relative section of Chapters II-IV.

Chapter I Summary

- *Metacognition* is our ability to reflect on our own thoughts and behaviours (Flavell, 1979).
- *Type-1* performance refers to an individual's performance in a cognitive task, while *type-2* performance corresponds to the second-order process of *monitoring* and *controlling* type-1 performance (Nelson, 1990).
- Trial-by-trial *confidence judgments* serve as a useful tool to measure an individual's metacognitive monitoring abilities by assessing their level of certainty regarding the accuracy of their responses. In the present thesis, the index of type-2 metacognitive efficiency (Mratio = type-2 sensitivity / type1-sensitivity) is used to statistically quantify how well confidence judgments track task accuracy, avoiding the confounding effects of confidence and type-1 biases (Fleming, 2017).
- *Metacognitive monitoring* is suggested to develop with age, especially during childhood (Destan et al., 2014; Rohwer et al., 2012), but its *developmental course* still remains unknown.
- Another hotly debated topic in the neuro-development of metacognition is whether this ability is supported by *domain general/specific mechanisms*. Hitherto, a gradual shift from domain specific to domain general mechanisms supporting metacognition has been suggested to occur during primary school years (Geurten et al., 2018; Lyons & Ghetti, 2010; Vo et al., 2014).
- The present study will focus on studying metacognitive monitoring in relation to reading acquisition during the first years of primary school.
- Reading acquisition is a complex process, requiring the development of cognitive and linguistic skills, which start developing from an early age (Georgiou et al., 2012; González-Valenzuela et al., 2016).
- Theories of reading development suggest that during reading instruction children transition from early reading stages, using decoding and letter-to-sound mappings to read, to more advanced reading stages, in which students mainly use larger grain orthographic representations and automatic visual word recognition to read (Ehri, 1995; Frith, 1985).

- The development of computational models of visual word recognition (e.g., DRC, PDP, MTM models) has helped us to understand the underlying processes which allow a reader to visually recognize a word. The interaction between orthography, phonology and semantics is key to all models (Coltheart, 2006).
- Previous research has highlighted the role of metacognition in reading and academic achievement, but has focused on reading comprehension (*meta-comprehension*) in later stages of reading development, when prerequisite skills of reading are already in place and typically developing children can read fluently (Garner, 1987; Jacobs & Paris, 1987; Roebers et al., 2009).
- The present study is focused on understanding the role of metacognition during the first years of reading instruction and with regard to the development of orthographic lexical processing, referring to individuals' orthographic knowledge and processing of whole word forms, which is essential for developing fluent reading in the first place. This is studied through the measurement of the following skills:

  (i) *Visual attention (VA) span*, a pre-orthographic processing skill allowing the parallel processing and identification of multiple visual elements under one eye fixation. The simultaneous processing of strings of letters has been shown to facilitate the formation of lexical orthographic traces in memory for future access (Bosse et al., 2007; Bosse & Valdois, 2009).

  (ii) *Orthographic lexical decision:* this relates to the efficiency of activating whole word orthographic representations in memory, and therefore, the quality of these representations, which partly relies on the efficiency of the abovementioned skills (Chetail, 2017; Ginestet et al., 2019).
- Here, we follow a within-subject longitudinal design to track students' task sensitivity and metacognitive efficiency in tasks assessing the above mentioned skills, a non-reading task for comparison, and also children's standardized reading ability, from Grade 1 to Grade 3.
- Understanding whether metacognitive monitoring is involved in the development of these skills can provide useful insights on how to support early readers in an educational context.

2. Chapter II: Setting up the bases for exploring the role of metacognition in primary school children: Characterising task performance in linguistic and nonlinguistic tasks.

2.1. Introduction

The present chapter aims to assess the relationship between type-1 task performance in the linguistic and non-linguistic tasks used in the present thesis, in order to set the bases for then exploring the functional role of metacognitive monitoring of the same tasks during the first years of primary school. Specifically, we provide estimates of abilities related to orthographic lexical processing and of reading performance skills and we explore the associations of these variables in the first years of primary school (Grade 1 and Grade 3 students). Moreover, we assess students' performance in a non-reading related task, in order to provide evidence that these associations are reading specific.

We aim to investigate: a) whether task sensitivity in tasks related to orthographic lexical processing and a non-linguistic task, unrelated to reading skills, is associated with students' performance in standardized reading tests (reading accuracy and reading time) within and across the two time points of the longitudinal study (Grades 1 and Grade 3), b) whether task performance in the same tasks changes over the course of the first two years of primary school, and c) whether students' performance in time point 1 correlates with their learning improvement in each task. Thereby, the present chapter will assess the following research questions:

2.1.1. Does early task sensitivity in experimental tasks assessing orthographic lexical processing correlate across them and with students' standardized reading performance within and across the first years of primary school?

As presented in the General Introduction, orthographic acquisition relates to the efficient processing of whole-word orthographic forms and plays a crucial role in the development of reading fluency (Adams, 1994; Stanovich, 1980). A critical prerequisite skill for mastering orthographic knowledge is one's visual attention span (VA span), which is related to the capacity to visually process in parallel multiple letters forming orthographic units. This skill has been shown to contribute both in single word identification skills and to predict students' current and later reading fluency (Besner et al., 2016; Valdois et al., 2019). Specifically, VA span has been found to strongly associate with students' reading time across the primary school years, while its' contribution to reading accuracy of any reading item (words and pseudowords) has been found to be stronger in Grade 1, and to restrict mostly to irregular words reading in later stages of primary school. Moreover, deficits in one's VA span have been linked to reduced reading accuracy and time (Valdois et al., 2003). Lexical orthographic knowledge has been considered to provide the link between the visual processing of printed words and reader's access to the phonology of those, which is considered essential for fluent reading (Stanovich, 1993).

If orthographic knowledge plays a crucial role in the development of strong reading abilities, as previously found in the literature, significant correlations are expected between: a) type-1 sensitivity in the tasks assessing lexical orthographic knowledge (i.e., orthographic lexical decision) and sensitivity in the task assessing its prerequisite skill (i.e., VA span task) and, b) type-1 sensitivity in these tasks and students' performance in the standardized tests measuring reading accuracy and reading time. Our hypothesis is that performance in the tasks assessing orthographic lexical knowledge and its prerequisites, will positively correlate across these tasks and with participants' standardized reading accuracy, and negatively correlate with participants' reading time within both time points of the study. Moreover, we expect that the contribution of VA span in reading accuracy will be larger in Grade 1 than in Grade 3, as previously shown in the literature (e.g., Ginestet et al., 2021; Valdois et al., 2019). On the contrary, we expect no significant correlations between the reading tasks and the non-reading related, emotion recognition task.

Finally, if VA span and lexical orthographic knowledge are future predictors of reading accuracy and time, we expect to see links between sensitivity in these tasks in Grade 1, and students' future standardized reading performance.

2.1.2. Does reading performance change significantly during early reading development?

Lexical orthographic knowledge has been suggested to show an accelerated development during the first years of primary school and reaches mastery around the fourth year of primary school (Adams, 1994; Stanovich, 1980). Moreover, individuals' VA span, a pronounced prerequisite of orthographic knowledge, has been found to significantly increase during the first years of primary school (Grade 1 to Grade 3, focus age groups in our study) and continue improving later on in development (Bosse & Valdois, 2009; Popa, 2020). Finally, reading performance has been found to significantly increase in older children compared to first graders', independently of the reading materials (word/pseudoword, regular/irregular items, see: Popa, 2020).

Here, we examine the development of these skills throughout the first years of primary school, in order to confirm that our data follow these patterns, before assessing the development of metacognition in the same tasks in the following Chapters of the thesis. Hence, we expect that students' task performance both in tasks assessing orthographic knowledge and its prerequisite skills, but also in reading accuracy in Grade 3 will be higher than in Grade1, after children have received 2 full years of reading instruction, and that reading time will reduce across the two time points of the study. Finally, we expect that task performance in the emotion recognition task will also exhibit improvements across time, as previously shown in the literature (Lawrence et al., 2015).

2.1.3. Does early reading performance predict improvements in reading across time?

In the beginning of primary school, when students start receiving formal instruction in reading, there is a big heterogeneity of reading skills in the classroom, with some children entering primary school having received extended reading instruction at home or at the kindergarten, while others have no experience in reading. Previous literature indicates that children who learn how to read later on than their peers, catch up by the age of 11 (Suggate et al., 2013). In the present section, we wanted to additionally examine how and if early performance in tasks related to orthographic lexical processing and reading ability uniquely contributes to longitudinal improvements (i.e., learning effect) in the same tasks. Moreover, we sought to determine whether predictors of orthographic lexical processing predict changes in standardized reading performance across the first years of primary school.

Our hypothesis is that, if less skilled readers in the beginning of reading acquisition learn more during the first two years of primary school to catch up with children who entered primary school with more advanced reading skills, they will be the ones showing greater longitudinal improvements in the reading-related tasks over time, and that variability in reading performance will reduce after 2 years of formal instruction in reading.

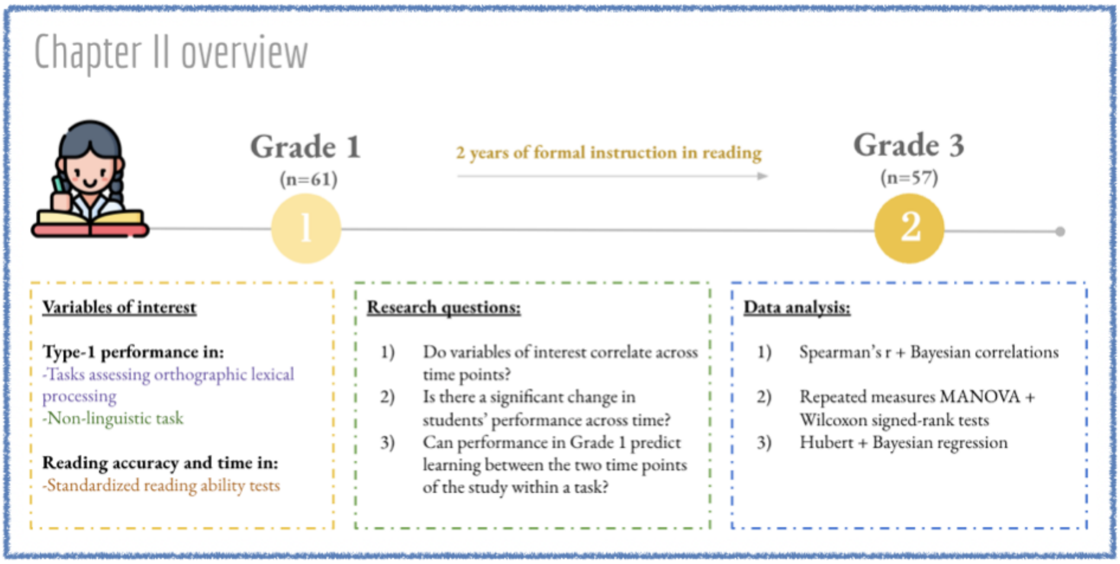


Figure 10 Schematic representation of the research design used in Chapter II.

2.2. Methods

2.2.1. Participants

As mentioned in the Chapter of Methodological Considerations, sixty-one children, native Spanish speakers, aged between 6 and 7 years (mean age (±SD): 6.59 ± 0.29, 28 girls), were assessed in the middle of Grade 1 and included in the data analysis of the first branch of the present longitudinal study. The second assessment point of this study occurred 2 years later, in Grade 3, when the sample consisted of fifty-seven children (mean age (±SD): 8.48 ± 0.34, 25 girls). The reduction of the sample was due to outliers identified in the first time point of the study, that were not re-assessed in Grade 3 or participants' change of school. A detailed description of participants' characteristics and process of recruitment can be found in Chapter Section 1.3.2.1.

2.2.2. Materials and Methods

Participants performed two linguistic tasks related to orthographic lexical processing; a VA span task, assessing the homonymous prerequisite skill of orthographic lexical processing and a lexical decision task, assessing children's orthographic lexical knowledge, and one non-linguistic, emotion recognition task, in order to assess any domain-specific effects. In addition to these tasks, participants were also administered two standardized reading tests (word and pseudoword reading) as well as a control task aimed at measuring their non-verbal reasoning abilities (non-verbal IQ). The details of all tasks administered and the assessment process can be found in the Chapter Section 1.3.2.3.

2.2.3. Data analysis

Prior to the main data analysis, participants performing at chance level in an experimental task were excluded from the analysis of the certain task (accuracy ≤ 0.5). Subsequently, the interquartile range (IQR) criterion was used to screen for outliers. Values that are two interquartile ranges larger or smaller from the median, were identified as outliers. Accuracy data were screened for outliers, and outliers were excluded separately in each experiment (see Table 2 for No of participants included in each task and time point after screening for outliers).

Next, as mentioned in the Chapter Section 1.3.2.4., task performance in each of the experimental tasks (VA span, lexical decision and emotion recognition task) was estimated as the type-1 task sensitivity (d-prime) using a signal detection theoretic framework for each time point of the study (Fleming, 2017). Moreover, for the tasks in which an adaptive staircase was used to adjust individual participants' accuracy at around 70% (see Chapter Section 1.3.2.3.), the mean stimulus presentation time was also calculated for each participant as an additional measure of type-1 task performance. Task performance in the standardized reading tests was calculated as: a) the average accuracy of words and pseudowords list reading (% items read correctly) and b) the average speed of words and pseudowords list reading (measured in sec). Finally, the difference between the estimates of each variable between the two time points of the study was calculated as a measure of performance change (i.e., learning effect) across time. Larger learning effects are indicated by a positive change in type-1 task sensitivity and reading accuracy variables and by a negative change in stimulus presentation and reading time variables.

Normality tests were then conducted in each variable of interest of the experimental tasks and the standardized reading tests. The measure of Skewness was used to evaluate normality. Considering that many values were moderately to highly skewed (Skewness > 0.5 or < -0.5), we elected to use non-parametric tests for the frequentist testing, when possible.

In order to investigate how type-1 task performance in all the experimental tasks relates across the tasks and to participants' performance in the standardized reading tasks (reading accuracy and reading time) in each time point of the study, but also with their difference across time, Spearman's r correlations were used. For each correlational analysis involving the different aims set up in the introduction, False Discovery Rate (FDR) was used for multiple comparison correction and participants' chronological age and intellectual ability were used as covariates.

Subsequently, we aimed at identifying possible differences in participants' type-1 task performance in the experimental tasks (d-prime, stimulus presentation time) and in their standardized reading performance (reading accuracy and reading time) across the two time points of the study (Grade 1 and Grade 3). To this end, one-way repeated measures MANOVAS (Multivariate Analysis of Variance) were applied for each variable measured in the experimental tasks (type-1 d-prime, mean stimulus presentation duration) using the SPSS software, Version 28.0. When a main effect of time was discovered, univariate post hoc analysis, Bonferroni correction was applied to find out significant pairwise differences. In case of violation of sphericity assumptions, Greenhouse-Geisser correction was used to correct the p-values. In case of violation of normality assumption, log transformation of the data was applied, and when normality was still violated in some conditions, post-hoc analyses were repeated on those using non-parametric Wilcoxon signed-rank tests to confirm the pattern of results. Wilcoxon-signed rank tests were also applied to compare reading accuracy and reading time across timepoints of the study.

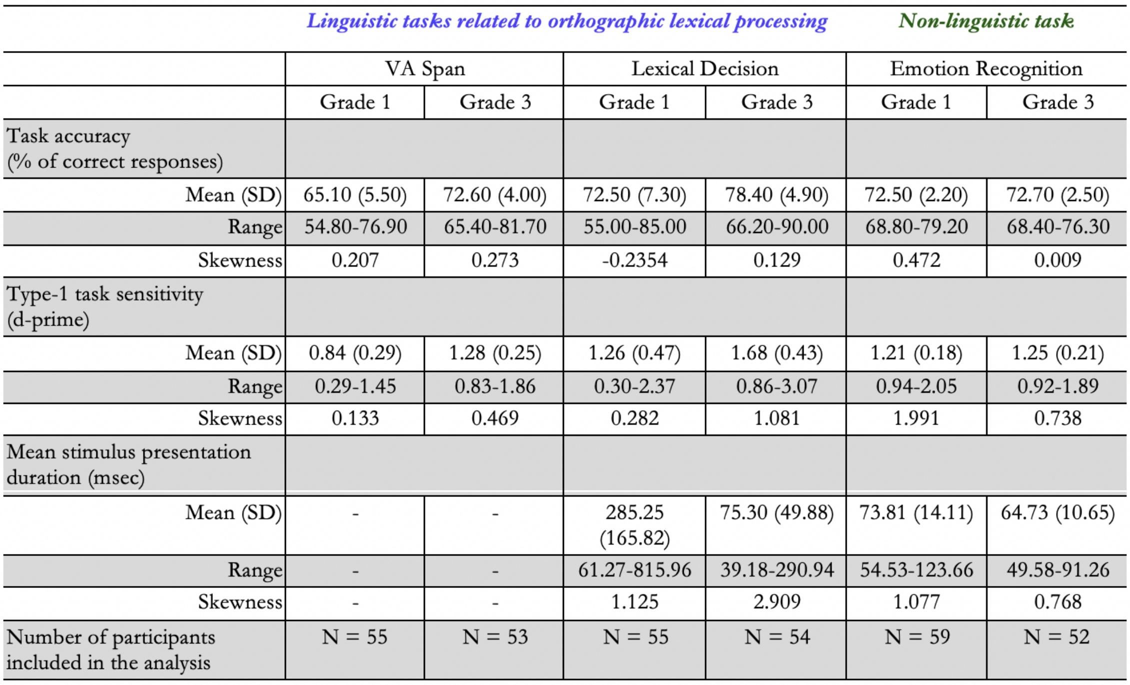
Finally, we investigated the longitudinal links between participants' task performance in a reading related task (VA span, lexical decision, standardized reading tests) and long-term performance improvement in the same task, using linear and Bayesian regression analyses. For the linear regressions, Huber robust regression was applied, which accounts for outliers. For the Bayesian regressions, a default prior of 0.354, as implemented in JASP software (van Doorn et al., 2020) was used and the Bayesian inclusion factor ($BF_{inclusion}$) was estimated for every predictor in the model. $BF_{inclusion}$ is calculated by dividing the prior odds of a model including a predictor of interest by the posterior odds (i.e., BF_{10}) excluding this predictor. When $BF_{inclusion} > 1$, it indicates that the model was improved by the addition of this specific predictor.

2.3. Results

2.3.1. Descriptives

Table 2 summarises the descriptive analysis for the measures used to assess type-1 performance (accuracy, d-prime, mean stimulus presentation duration) in the different experimental tasks and Table 3 includes the descriptive analysis for participants' performance in the standardized reading tests (reading accuracy and reading time).

Table 2 Descriptive statistics (Mean Score (SD), Range, Skewness) for the different measures of participants' type-1 performance (accuracy, d-prime, mean stimulus presentation duration).



	Linguistic tasks related to orthographic lexical processing				*Non-linguistic task*	
	VA Span		Lexical Decision		Emotion Recognition	
	Grade 1	Grade 3	Grade 1	Grade 3	Grade 1	Grade 3
Task accuracy (% of correct responses)						
Mean (SD)	65.10 (5.50)	72.60 (4.00)	72.50 (7.30)	78.40 (4.90)	72.50 (2.20)	72.70 (2.50)
Range	54.80-76.90	65.40-81.70	55.00-85.00	66.20-90.00	68.80-79.20	68.40-76.30
Skewness	0.207	0.273	-0.2354	0.129	0.472	0.009
Type-1 task sensitivity (d-prime)						
Mean (SD)	0.84 (0.29)	1.28 (0.25)	1.26 (0.47)	1.68 (0.43)	1.21 (0.18)	1.25 (0.21)
Range	0.29-1.45	0.83-1.86	0.30-2.37	0.86-3.07	0.94-2.05	0.92-1.89
Skewness	0.133	0.469	0.282	1.081	1.991	0.738
Mean stimulus presentation duration (msec)						
Mean (SD)	-	-	285.25 (165.82)	75.30 (49.88)	73.81 (14.11)	64.73 (10.65)
Range	-	-	61.27-815.96	39.18-290.94	54.53-123.66	49.58-91.26
Skewness	-	-	1.125	2.909	1.077	0.768
Number of participants included in the analysis	N = 55	N = 53	N = 55	N = 54	N = 59	N = 52

Table 3 Descriptive statistics (Mean Score (SD), Range, Skewness) for the different measures of participants' standardized reading performance (reading accuracy, reading time).

	Standardised reading tasks	
	Grade 1	Grade 3
Reading accuracy (% of correct responses)		
Mean (SD)	0.71 (0.27)	0.92 (0.07)
Range	0.03-0.98	0.73-1.00
Skewness	-1.135	-1.135
Reading time (sec)		
Mean (SD)	197.23 (113.67)	63.10 (17.45)
Range	53.00-598.00	39.00-118.00
Skewness	1.249	1.031

2.3.2. Correlations between type-1 task performance in the experimental tasks and students' performance in standardized reading tests in Grade 1 and Grade 3

First, correlations between participants' type-1 task sensitivity across the experimental tasks and their performance on the standardized reading tests (reading accuracy and reading time) were performed within and across time points of the studies. Second, correlation analyses were run between student's task performance in the experimental and the standardized reading tasks in Grade 1 and learning improvements (difference in variables) between Grade 1 and Grade 3.

We expected (i) participants with higher type-1 task sensitivity in tasks assessing orthographic lexical processing (VA span, lexical decision) to exhibit higher performance on the standardized reading tests (e.g., Ginestet et al., 2021; Valdois et al., 2019) within both time points of the study and (ii) participants with higher type-1 task performance in VA span and the lexical decision task in Grade 1 to exhibit better performance across reading-related tasks in Grade 3, but smaller changes in reading performance across time points.

Type-1 task sensitivity and standardized reading skills in Grade 1 and Grade 3

First, strong associations were shown in students' standardized reading ability measures (reading accuracy and reading time) both within and across time points (all ps < 0.05). Specifically, both students' reading accuracy and reading time positively correlated across Grade 1 and Grade 3 ($p < 0.001$), while reading accuracy was negatively associated with reading time (higher accuracy, less time) both within and across timepoints. The same pattern of results was repeated in the associations between standardized reading measures and type-1 task performance (type-1 d-prime and stimulus presentation duration measures) in the lexical decision task within and across time points of the study (all ps < 0.05, see Table 4). Moreover, significant or marginally significant associations were found between participants' type-1 task sensitivity in the VA span task in Grade 1 and the standardized reading measures (VA span d-prime-reading accuracy: $r = 0.299$, $p = 0.087$, VA span d-prime-reading time: $r = -0.360$, $p = 0.028$), for which Bayes factor provided moderate evidence towards the alternative hypothesis (all $BF_{10} > 3$, see Table 4). Bayes factor also provided anecdotal evidence towards the alternative hypothesis in the associations between VA span type-1 task sensitivity in Grade 1 and type-1 task performance in the lexical decision task in Grade 1, but also participants' reading accuracy in Grade 3 (all $BF_{10} > 1$, see Table 4). This pattern of results was not observed in Grade 3 (all ps > 0.23, all $BF_{10} < 0.51$, see Table 4). Finally, type-1 performance on the non-linguistic task did not correlate with any of our reading-related tasks in any of the time points of the study (all ps > 0.05, see Table 4 for details).

Type-1 task performance in Grade 1 and learning improvements between Grade 1 and Grade 3

Strong negative correlations were found between participants' task performance in all the experimental tasks and the standardized reading tests in Grade 1, and their learning improvements across time points within the same task (all ps < 0.001, see Table 5). Here, it is important to mention that in variables measuring time (i.e. reading time and mean stimulus presentation duration in the lexical decision and the emotion recognition task), lower values indicate better performance (faster in time) and negative values in the difference between the time points of the study in these variables, indicate higher learning improvements. Hence, here negative correlations between reading time in the standardized tests and stimulus presentation duration in the lexical decision and the emotion recognition task, and their respective learning improvements, indicate that slower students were the ones who improved more in these same variables. Similarly, negative correlations in reading accuracy in the standardized reading test and task sensitivity in all the experimental tasks (d-prime) and their respective improvements, indicate that students with lower performance in Grade 1 were the ones who improved more in these variables.

Next, standardized reading accuracy in Grade 1 was positively associated both with the learning improvements in standardized reading time and with the difference in the mean stimulus presentation duration on the lexical decision task across time points (all ps < 0.001, see Table 5), indicating again that students showing lower reading accuracy in Grade 1 are the ones improving more in accuracy and in processing speed, namely, reducing the mean stimulus presentation duration in the lexical decision task across the two time points. The same pattern of results was repeated when associating type-1 task sensitivity (d-prime) in the lexical decision task and learning in the rest of the variables (all ps≤ 0.005, see Table 5). Next, standardized reading time in Grade 1 was found to positively correlate with learning improvements in reading accuracy, and negatively correlate with learning improvements in the mean stimulus presentation duration in the lexical decision task (slower reading in Grade 1, more learning over time). A similar pattern of results was found when correlating task performance in the lexical decision task measured as the stimulus presentation duration and learning improvements in the same variables (all ps < 0.001, see Table 5).

Table 4 Spearman's correlations between type-1 task performance in the experimental tasks and: a) students' standardized reading ability (reading accuracy and speed) and b) type-1 task performance in the rest of the experimental tasks (*p < 0.05, **p < 0.01, ***p < 0.001, FDR corrected) within and across Grade 1 and Grade 3. Correlations were controlled for participant's age and intellectual ability (non-verbal IQ, Matrices). Significant correlations are noted in bold font.

	Standardised reading tasks				***Linguistic tasks related to orthographic lexical processing***						***Non-linguistic task***			
	Reading Accuracy (%correct)		Reading Time (sec)		VA Span d-prime		Lexical Decision d-prime		Lexical Decision stimulus duration (msec)		Emotion Recognition d-prime		Emotion Recognition stimulus duration (msec)	
	Grade 1	**Grade 3**	**Grade 1**	**Grade 3**	**Grade 1**	**Grade 3**	**Grade 1**	**Grade 3**	**Grade 1**	**Grade 3**	**Grade 1**	**Grade 3**	**Grade 1**	**Grade 3**
Reading Accuracy (%correct) Grade 1	-	**r = 0.636 p < 0.001 BF_{10}>100**	**r = -0.703 p < 0.001 BF_{10} >100**	**r = -0.432 p = 0.009 BF_{10}=3.519**	r = 0.299 p = 0.087 BF_{10}=4.110	r = 0.216 p = 0.302 BF_{10}=0.230	**r = 0.528 p = 0.001 BF_{10} >100**	**r = 0.380 p = 0.025 BF_{10}=2.015**	**r = -0.704 p <0.001 BF_{10} >100**	**r = -0.493 p = 0.002 BF_{10}=1.768**	r = -0.049 p = 0.845 BF_{10}=0.164	r = -0.032 p = 0.877 BF_{10}=0.180	r = 0.122 p = 0.584 BF_{10}=0.167	r = -0.091 p = 0.699 BF_{10}=0.371
Reading Accuracy (%correct) Grade 3		-	**r = -0.432 p = 0.009 BF_{10}=8.006**	**r = -0.355 p = 0.033 BF_{10}=0.991**	r = 0.169 p = 0.463 BF_{10}=1.363	r = 0.247 p = 0.232 BF_{10}=0.449	**r = 0.438 p = 0.010 BF_{10}=3.974**	**r = 0.465 p = 0.005 BF_{10}=11.617**	**r = -0.381 p = 0.028 BF_{10}=1.638**	**r = -0.505 p =0.002 BF_{10}=5.203**	r = 0.026 p = 0.877 BF_{10}=0.176	r = -0.046 p = 0.845 BF_{10}=0.186	r = -0.126 p = 0.593 BF_{10}=0.194	r = -0.159 p = 0.500 BF_{10}=0.232
Reading Time (sec) Grade 1			-	**r = 0.644 p < 0.001 BF_{10}>100**	**r = -0.360 p = 0.028 BF_{10}=42.442**	r = -0.068 p = 0.790 BF_{10}=0.177	**r = -0.419 p = 0.010 BF_{10}=7.632**	**r = -0.384 p = 0.023 BF_{10}=2.373**	**r = 0.728 p < 0.001 BF_{10}>100**	**r = 0.491 p = 0.002 BF_{10}=0.699**	r = 0.129 p = 0.563 BF_{10}=0.235	r = 0.044 p = 0.845 BF_{10}=0.177	r = -0.120 p = 0.584 BF_{10}=0.293	r = -0.091 p = 0.699 BF_{10}=0.203
Reading Time (sec) Grade 3				-	r = -0.228 p = 0.279 BF_{10}=0.742	r = -0.001 p = 0.997 BF_{10}=0.192	**r = -0.381 p = 0.028 BF_{10}=2.369**	**r = -0.393 p = 0.022 BF_{10}=10.887**	**r = 0.540 p = 0.001 BF_{10}=15.164**	**r = 0.586 p < 0.001 BF_{10}=17.488**	r = 0.095 p = 0.699 BF_{10}=0.187	r = 0.198 p = 0.368 BF_{10}=0.238	r = -0.042 p = 0.845 BF_{10}=0.204	r = -0.136 p = 0.584 BF_{10}=0.204
VA Span d-prime Grade 1					-	r = 0.120 p = 0.656 BF_{10}=0.274	r = 0.297 p = 0.116 BF_{10}=1.061	r = 0.191 p = 0.385 BF_{10}=0.210	r = -0.269 p = 0.172 BF_{10}=1.823	r = -0.234 p = 0.270 BF_{10}=0.560	r = 0.077 p = 0.740 BF_{10}=0.189	r = -0.109 p = 0.669 BF_{10}=0.188	r = 0.237 p = 0.229 BF_{10}=0.212	r = -0.113 p = 0.656 BF_{10}=0.321
VA Span d-prime Grade 3						-	r = 0.204 p = 0.368 BF_{10}=0.314	r = 0.068 p = 0.793 BF_{10}=0.230	r = -0.163 p = 0.500 BF_{10}=0.194	r = 0.001 p = 0.997 BF_{10}=0.513	r = 0.178 p = 0.419 BF_{10}=0.534	r = 0.278 p = 0.172 BF_{10}=0.557	r = -0.028 p = 0.877 BF_{10}=0.190	r = -0.174 p = 0.460 BF_{10}=0.263
Lexical Decision d-prime Grade 1							-	**r = 0.412 p = 0.019 BF_{10}=2.858**	**r = -0.415 p = 0.010 BF_{10}=17.616**	**r = -0.414 p = 0.019 BF_{10}=2.719**	r = 0.187 p = 0.368 BF_{10}=0.548	r = 0.048 p = 0.845 BF_{10}=0.199	r = -0.088 p = 0.699 BF_{10}=0.186	r = -0.222 p = 0.331 BF_{10}=1.011
Lexical Decision d-prime Grade 3								-	**r = -0.357 p = 0.045 BF_{10}=0.434**	**r = -0.54 p = 0.001 BF_{10}=1.088**	r = -0.146 p = 0.532 BF_{10}=0.176	r = -0.263 p = 0.212 BF_{10}=0.337	r = -0.027 p = 0.877 BF_{10}=0.194	r = -0.027 p = 0.877 BF_{10}=0.194
Lexical Decision stimulus duration (msec) Grade 1									-	**r = 0.553 p = 0.001 BF_{10}=1.705**	r = -0.083 p = 0.718 BF_{10}=0.397	r = -0.051 p = 0.845 BF_{10}=0.209	r = -0.218 p = 0.279 BF_{10}=0.486	r = -0.119 p = 0.656 BF_{10}=0.194
Lexical Decision stimulus duration (msec) Grade 3										-	r = 0.039 p = 0.856 BF_{10}=0.175	r = 0.072 p = 0.790 BF_{10}=0.303	r = -0.181 p = 0.403 BF_{10}=0.510	r = 0.094 p = 0.699 BF_{10}=0.619
Emotion Recognition d-prime Grade 1											-	r = -0.112 p = 0.656 BF_{10}=0.398	r = -0.182 p = 0.368 BF_{10}=0.167	r = -0.056 p = 0.845 BF_{10}=0.178
Emotion Recognition d-prime Grade 3												-	r = 0.110 p = 0.656 BF_{10}=0.182	**r = -0.376 p = 0.028 BF_{10}=5.344**
Emotion Recognition stimulus duration (msec) Grade 1													-	r = 0.050 p = 0.845 BF_{10}=0.248

Table 5 Spearman's correlations between task performance in the experimental tasks and the standardized reading tasks in Grade 1 and longitudinal changes of the same variables across the two timepoints (difference in performance between Grade 1 and Grade 3) (*p < 0.05, **p < 0.01, ***p < 0.001, FDR corrected). Correlations were controlled for participant's age and intellectual ability (non-verbal IQ, Matrices). Significant correlations are noted in bold font.

	Standardised reading tasks		*Linguistic tasks related to orthographic lexical processing*			*Non-linguistic task*	
	Reading Accuracy (%correct) Δ (Grade 3-Grade 1)	Reading Time (sec) Δ (Grade 3-Grade 1)	VA Span d-prime Δ (Grade 3-Grade 1)	Lexical Decision d-prime Δ (Grade 3-Grade 1)	Lexical Decision Mean stimulus duration (msec) Δ (Grade 3-Grade 1)	Emotion Recognition d-prime Δ (Grade 3-Grade 1)	Emotion Recognition Mean stimulus duration (msec) Δ (Grade 3-Grade 1)
Reading Accuracy (%correct) Grade 1	**r = -0.956** **p < 0.001** **BF$_{10}$ > 100**	**r = 0.755** **p < 0.001** **BF$_{10}$ > 100**	r = -0.010 p = 0.969 BF$_{10}$ = 0.312	r = -0.192 p = 0.455 BF$_{10}$ = 0.983	**r = 0.606** **p < 0.001** **BF$_{10}$ > 100**	r = -0.005 p = 0.985 BF$_{10}$ = 0.182	r = -0.143 p = 0.674 BF$_{10}$ = 0.272
Reading Time (sec) Grade 1	**r = 0.741** **p < 0.001** **BF$_{10}$ > 100**	**r = -0.980** **p < 0.001** **BF$_{10}$ > 100**	r = 0.099 p = 0.799 BF$_{10}$ = 1.381	r = 0.049 p = 0.829 BF$_{10}$ = 0.241	**r = -0.645** **p < 0.001** **BF$_{10}$ > 100**	r = -0.030 p = 0.911 BF$_{10}$ = 0.223	r =-0.081 p = 0.799 BF$_{10}$ = 0.206
VA Span d-prime Grade 1	r = -0.249 p = 0.241 BF$_{10}$ = 1.147	r = 0.304 p = 0.118 BF$_{10}$ = 18.222	**r = -0.679** **p < 0.001** **BF$_{10}$ > 100**	r = -0.145 p = 0.684 BF$_{10}$ = 0.327	r = 0.097 p = 0.799 BF$_{10}$ = 0.362	r = -0.103 p = 0.799 BF$_{10}$ = 0.194	r = -0.171 p = 0.491 BF$_{10}$ = 0.213
Lexical Decision d-prime Grade 1	**r = -0.464** **p = 0.004** **BF$_{10}$ > 100**	**r = 0.456** **p = 0.005** **BF$_{10}$ = 8.062**	r = -0.016 p = 0.965 BF$_{10}$ = 0.196	**r = -0.538** **p < 0.001** **BF$_{10}$ > 100**	r = 0.322 p = 0.096 BF$_{10}$ =1.631	r = -0.096 p = 0.799 BF$_{10}$ =0.266	r = -0.055 p = 0.828 BF$_{10}$ = 0.256
Lexical Decision Mean stimulus duration (msec) Grade 1	**r = 0.677** **p < 0.001** **BF$_{10}$ > 100**	**r = -0.711** **p < 0.001** **BF$_{10}$ > 100**	r = -0.072 p = 0.804 BF$_{10}$ = 0.199	r = 0.118 p = 0.761 BF$_{10}$ = 0.384	**r = -0.912** **p < 0.001** **BF$_{10}$ > 100**	r = 0.068 p = 0.804 BF$_{10}$ = 0.210	r = 0.084 p = 0.799 BF$_{10}$ = 0.259
Emotion Recognition d-prime Grade 1	r = -0.002 p = 0.990 BF$_{10}$ = 0.179	r = -0.042 p = 0.848 BF$_{10}$ = 0.355	r = 0.052 p = 0.828 BF$_{10}$ = 0.242	r = -0.282 p = 0.172 BF$_{10}$ = 0.674	r = 0.108 p = 0.799 BF$_{10}$ = 0.292	**r = -0.641** **p < 0.001** **BF$_{10}$ > 100**	r =0.129 p = 0.703 BF$_{10}$ = 0.199
Emotion Recognition Mean stimulus duration (msec) Grade 1	r = -0.066 p = 0.804 BF$_{10}$ = 0.184	r = -0.071 p = 0.800 BF$_{10}$ = 0.185	r = -0.147 p = 0.647 BF$_{10}$ = 0.186	r = 0.092 p = 0.799 BF$_{10}$ = 0.328	r = -0.093 p = 0.799 BF$_{10}$ = 0.185	r = 0.208 p = 0.394 BF$_{10}$ = 0.197	**r = -0.731** **p < 0.001** **BF$_{10}$ > 100**

2.3.3. Comparing participants' type-1 performance in the experimental tasks and task performance in standardized reading tests across time

Type-1 task sensitivity (d-prime)

We first compared the task performance in the three experimental tasks across time measured as the log-transformed type-1 d-prime task sensitivity. A one-way repeated-measures MANOVA with time (Grade 1, Grade 3) as a within-subject factor on participants' type-1 task sensitivity in the three experimental tasks (VA span, lexical decision, emotion recognition), revealed a main effect of the time point of the study ($F(3,34) = 24.447$, $p < 0.001$, $\eta_p^2 = 0.683$). In order to confirm that the time effect applied to all the tasks, univariate post-hoc analysis was performed, which revealed that task performance showed an increase from Grade1 to Grade 3 within the VA span and the lexical decision task (all ps < 0.001), but showed no difference within the emotion recognition task ($p = 0.260$, Bonferroni-corrected). The post hoc analysis was repeated using pair-to-pair non-parametric Wilcoxon signed-rank tests due to the fact that some of the log transformed variables remained not normally distributed. The same pattern of results was reproduced.

Next, intersubject correlations showed that type-1 task sensitivity did not correlate across time points in any of the tasks (all ps > 0.05). A marginal positive correlation was indicated in the lexical decision task ($r = 0.412$, $p = 0.067$).

Stimulus presentation duration

As mentioned in Chapter Section 1.3.2.3., in the lexical decision and the emotion recognition task, an online staircase was used to adjust the stimulus presentation time of the stimuli in order to achieve overall accuracy of around 70% for each participant. Here, we compared the mean stimulus presentation duration in these two experimental tasks across the time points of the study. These variables were not normally distributed even after log transformation was applied in the data, so it was decided to perform a parametric one-way MANOVA in the non-transformed data, followed by non-parametric Wilcoxon signed-rank pair-to-pair tests.

A one-way repeated-measures MANOVA with time (Grade 1, Grade 3) as a within-subject factor on participants' mean stimulus presentation duration in lexical decision and emotion recognition task, revealed a main effect of the time point of the study ($F(2,40) = 53.341$, $p < 0.001$, $\eta_p^2 = 0.727$). In order to confirm that the time effect applied to all the tasks, univariate post-hoc analysis was performed, which revealed that stimulus presentation duration showed an increase from Grade1 to Grade 3 within the lexical decision and the emotion recognition task (all ps < 0.003, Bonferroni-corrected). The post-hoc analysis was repeated using Wilcoxon signed-rank tests and the same pattern of results was reproduced.

Standardized reading measures (reading accuracy and reading time)

Finally, we compared participants' standardized reading performance (reading accuracy and reading time) across the two time points of the study. Wilcoxon signed-rank tests revealed a significant increase in participants' reading accuracy across time points ($W = 15.00$, $p < 0.001$) and a significant decrease in participants' reading time ($W = 1431.00$, $p < 0.001$) (see Figure 11c & 11d).

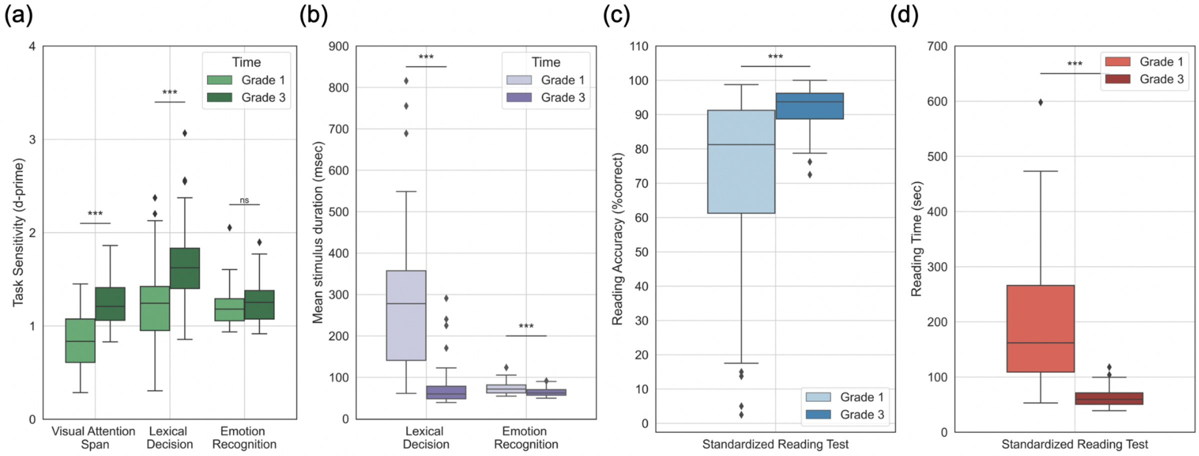


Figure 11 (a) Task sensitivity (d-prime) in the experimental tasks, (b) mean stimulus presentation duration (msec) in tasks utilizing an adaptive staircase, (c) standardized reading accuracy (% correct) and (d) standardized reading time (sec) in Grade 1 and Grade 3 of the longitudinal study. Significance is indicated for the effect of the time point of the study (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, Bonferroni correction).

In Figure 11c and 11d, one can notice that inter-individual variability in the standardized reading measures (indicated both by the vertical distance between the lowest and the highest value and the interquartile range of the boxplots), was lower in Grade 3 than in Grade 1. Low variability indicates that values in a sample are more consistent.

2.3.4. Investigating the relationship between participants' early reading performance and its developmental changes between Grade 1 and 3

We first examined whether task performance of students in Grade 1 in the reading-related tasks predicts longitudinal changes on their performance on the same task across the two time points of the study. Next, we tested whether task performance in tasks assessing orthographic lexical processing (VA span, lexical decision) predict changes in task performance in standardized reading tests and, finally, whether task performance in the VA span task, a predictor of orthographic lexical processing, can predict changes in the lexical decision task reflecting orthographic knowledge.

Regression analyses were performed to predict these performance changes, measured as the difference in task performance between Grade 1 and Grade 3 within each task, using task performance, age and IQ at Grade 1 as predictors. Analyses performed using a robust to outliers regression model (Hubert regression) showed that:

a) task performance in all reading-related tasks in Grade 1 was a significant negative predictor of longitudinal changes in the same variable within all tasks (all ps < 0.05, see Table 6, Figure 12 a-d).

b) task performance in the lexical decision task significantly and negatively predicted changes in reading accuracy and reading time in the standardized reading tests across time (all ps < 0.05, see Table 7, Figure 12 g-h), while performance in the VA span task significantly predicted changes in reading time ($p = 0.018$, see Figure 12e), and only marginally in reading accuracy ($p = 0.067$, see Figure 12f).

c) task performance in the VA span task did not predict changes in task performance in the lexical decision task across time ($p > 0.05$, see Table 8, Figure 12i).

The Bayesian regression models confirmed the above mentioned results (where ps < 0.05, all $BF_{inclusions} > 1$).

Table 6 Regression analysis (Hubert regression) of participants' longitudinal changes in task performance between the two time points of the longitudinal study in a) the VA Span task (n=47), b) the lexical decision task(n=48), c) the standardized reading tests (n=53), with task performance, age and non-verbal IQ at Grade 1 as predictors. $BF_{inclusion}$ factor represents the change from prior to posterior probabilities of a model when a predictor is added in the equation ($BF_{inclusion} > 1$ indicates that the predictor improves the model).

	d-prime changes in VA Span task (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
VA Span d-prime (Grade 1)	-0.874	-6.317	<0.001	>100	>100
Age	-0.002	-0.011	0.991	0.269	1.000
Non-verbal IQ	-0.009	-0.796	0.434	0.335	1.000

	d-prime changes in Lexical Decision task (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Lexical Decision d-prime (Grade 1)	-0.674	-6.006	<0.001	>100	>100
Age	0.154	0.864	0.390	0.612	1.000
Non-verbal IQ	-0.018	-1.088	0.276	0.951	1.000

	Reading Accuracy (%correct) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Reading Accuracy (%correct) (Grade 1)	-0.841	-31.725	<0.001	>100	>100
Age	0.025	1.055	0.298	0.105	1.000
Non-verbal IQ	0.002	0.769	0.441	0.075	1.000

	Reading Time (sec) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Reading Time (sec) (Grade 1)	-0.896	-50.714	<0.001	>100	>100
Age	-0.248	-1.583	0.118	0.112	1.000
Non-verbal IQ	0.403	0.771	0.444	0.044	1.000

Table 7 Regression analysis (Hubert regression) of participants' longitudinal changes in task performance in the standardized reading tests (reading accuracy, reading time) between the two time points of the longitudinal study, with task performance in a) the VA span task (n=47), b) the lexical decision task (n=48), age and non-verbal IQ at Grade 1 as predictors. $BF_{inclusion}$ factor represents the change from prior to posterior probabilities of a model when a predictor is added in the equation ($BF_{inclusion} > 1$ indicates that the predictor improves the model).

	Reading Accuracy (%correct) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
VA Span d-prime (Grade 1)	-0.177	-1.871	0.067	0.875	1.190
Age	-0.015	-0.173	0.863	0.360	1.000
Non-verbal IQ	-0.007	-0.859	0.392	0.548	1.00

	Reading Time (sec) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
VA Span d-prime (Grade 1)	117.176	2.500	0.018	8.704	10.115
Age	-17.974	-0.404	0.694	0.478	1.000
Non-verbal IQ	3.853	0.949	0.352	0.519	1.000

	Reading Accuracy (%correct) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Lexical Decision d-prime (Grade 1)	-0.198	-3.737	<0.001	>100	>100
Age	-0.036	-0.432	0.668	0.472	1.000
Non-verbal IQ	-0.003	-0.435	0.665	0.435	1.000

	Reading Time (sec) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Lexical Decision d-prime (Grade 1)	81.433	3.443	0.001	5.321	11.453
Age	37.102	0.982	0.336	0.668	1.000
Non-verbal IQ	3.181	0.953	0.346	0.515	1.000

Table 8 Regression analysis (Hubert regression) of participants' longitudinal changes in task performance in the lexical decision task between the two time points of the longitudinal study, with task performance in the VA span task (n=47), age and non-verbal IQ at Grade 1 as predictors. $BF_{inclusion}$ factor represents the change from prior to posterior probabilities of a model when a predictor is added in the equation ($BF_{inclusion} > 1$ indicates that the predictor improves the model).

	d-prime changes in Lexical Decision task (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
VA Span d-prime (Grade 1)	-0.279	-0.998	0.306	0.438	0.581
Age	0.206	0.777	0.448	0.405	1.000
Non-verbal IQ	-0.033	-1.361	0.171	0.795	1.000

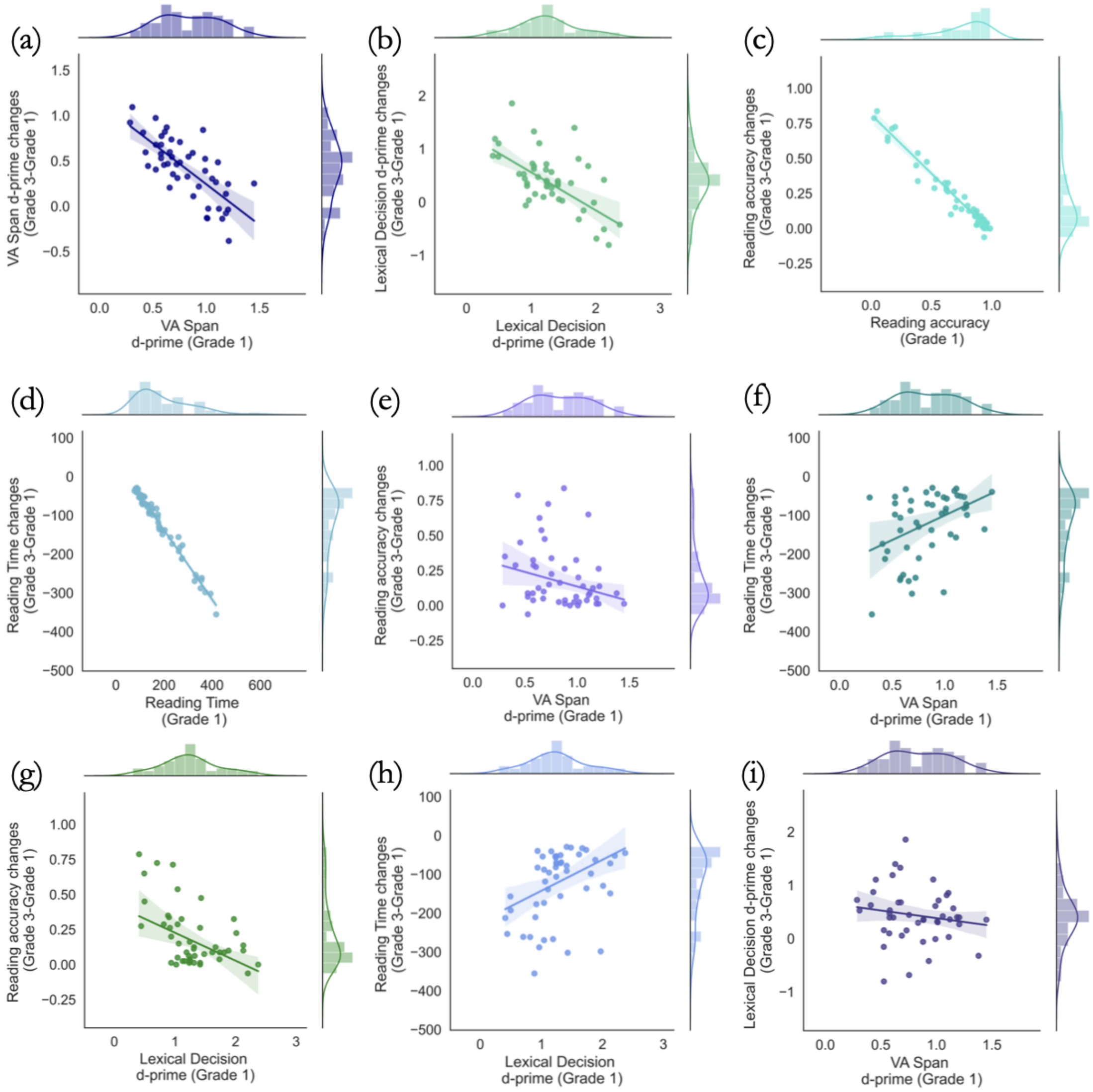


Figure 12 Linear relationship between longitudinal changes in performance in the reading-related tasks across time and students' performance in Grade 1 (a-d: within reading tasks, e-f: across reading tasks).

2.4. Discussion

The present chapter had the following purposes; first, we assessed the correlations between type-1 performance in the experimental tasks related to orthographic lexical processing and the non-linguistic task, with students' task performance in the standardized reading tests, within and across the two time points of the longitudinal study (Grade 1 and Grade 3). Second, we aimed at investigating if and how performance in all reading-related tasks changes across time. Finally, we sought to examine whether early performance in a reading task can predict learning effects in the same task across time and also whether early performance in tasks assessing predictors of orthographic lexical processing can predict learning effects in tasks assessing orthographic knowledge and standardized reading performance. In the following sections we discuss our findings for each research question separately.

Does early task sensitivity in experimental tasks assessing orthographic lexical processing correlate across them and with students' standardized reading performance across time?

Our first hypothesis was that participants' task performance in the tasks assessing orthographic lexical processing, but not in the non-linguistic task, will correlate to their standardized reading ability (reading accuracy and reading time) in both time points of the study.

First, we found that both in Grade 1 and Grade 3 students who exhibited higher task performance in the lexical decision task, reflecting orthographic knowledge, performed better in the standardized reading tasks, both in terms of reading accuracy and reading time, within time points. These results are in line with our expectations and follow the previous literature indicating that the lexical decision task significantly predicts individuals' word identification skills (Katz et al., 2012), and that orthographic knowledge, reflected by students' performance in the same task contributes to reading performance (Zarić et al., 2021).

Next, in Grade 1, students who showed higher task sensitivity in the VA span task, a predictor of orthographic lexical processing, were the ones exhibiting significantly lower reading time (faster reading) and marginally significant reading accuracy (see Table 4). Bayesian analysis revealed anecdotal evidence towards the alternative hypothesis ($BF_{10} > 1$) for the correlation between the tasks assessing orthographic lexical processing (VA span, lexical decision). However, in

Grade 3, no correlation was found between participants' VA span skills and the rest of the reading-related skills. Below we discuss possible interpretations of this finding.

VA span has been suggested to contribute to reading development from the early stages of reading acquisition and throughout the primary school years (Bosse & Valdois, 2009). Bosse and Valdois (2009), highlighted that the relationship between VA span skills was stronger in the beginning of primary school, a result borne out by the findings of the present study as well. This research group additionally found that from Grade 3 and on, VA span specifically contributes to the reading accuracy of irregular words, and its effect on regular word and pseudoword reading decreases over time (Bosse & Valdois, 2009). However, in the same study researchers reported that VA span significantly correlated with participants' reading time across primary school years (measured as the average time to read a word/pseudoword item).

In our study, no irregular words were used in order to verify the above mentioned findings regarding participants' reading accuracy, and no correlation was found between VA span in Grade 3 and standardized reading performance (accuracy and time) in regular words and pseudowords. These results may suggest that, in Grade 3, VA span skills could be less useful for readers at this stage of reading development as students rely more on lexical knowledge and sight word reading.

Moreover, a difference between our study and Bosse and Valdoid's study lays on the type of task used to assess VA span skills, including 4-letter string sequences instead of 5-letter strings respectively, and naming of the target letter in the sequence, instead of reporting presence or absence of the target letter in the sequence. These differences may have rendered, in our study, the task as easy to process by students in Grade 3, as an individual's VA span has been found to increase with age (Bosse & Valdois, 2009).

Finally, as expected, we found no correlations between students' task performance in the non-linguistic, emotion recognition task, and their standardized reading ability at any time point of the study.

Does reading performance change significantly during early reading development?

We hypothesized that students' performance both in experimental tasks assessing orthographic lexical processing and in standardized reading related tasks would increase across time in the first years of primary school. Our results demonstrated highly significant improvements between Grade

1 and Grade 3 in all tested variables related to reading (all ps < 0.001, see Figure 11). As mentioned in Chapter Section 2.1.2., predictors and markers of orthographic knowledge, as well as standardized reading performance, show an accelerated improvement during the first 3 years of primary school (Bosse & Valdois, 2009; Popa, 2020), which is reflected by our results. We also tested whether students' performance improves across time in a non-linguistic, emotion recognition task. We predicted that students will also exhibit improvements in this task, as emotion recognition skill has been suggested to naturally improve during development (Lawrence et al., 2015). We found no significant difference in students' task sensitivity (d-prime) in the emotion recognition task, which can be due to the fact that a staircase was used to adjust accuracy in this task, but a highly significant decrease in stimulus presentation time was noted across time, reflecting student's improvements across time in processing speed related to this skill.

Does early reading performance predict improvements in reading across time?

We predicted that: a) students' performance in the reading related tasks in Grade 1 will negatively correlate and uniquely predict the learning improvements of the students between the two time points of the study within tasks and b) student's performance in the tasks assessing orthographic lexical processing in Grade 1 will predict learning improvements in the standardized reading tests across time.

First, significant negative correlations were observed between students' task performance in all experimental and standardized reading tasks in Grade 1 and their learning improvements across time, quantified as the difference in these variables between Grade 1 and Grade 3. Regression analyses showed that early task performance predicted improvements in these tasks across time, independently of students' chronological age and non-verbal intellectual ability.

Second, a similar pattern of results was found when correlating students' task sensitivity in the lexical decision task in Grade 1 with students' learning improvements in the standardized reading ability tasks, both in reading accuracy and reading time measures. These correlations were not found to be significant in frequentist statistics in the case of the VA span task in Grade 1 and improvements in the standardized reading tasks, however Bayesian statistics provided supportive evidence towards the alternative hypothesis both for the reading accuracy and the reading time measures (see Table 5). Regression analyses revealed that early task performance in both tasks assessing orthographic lexical processing (VA span, lexical decision) significantly or marginally significantly predicted students' learning improvements in standardized reading accuracy and reading time across the two time points of the study (see Table 7).

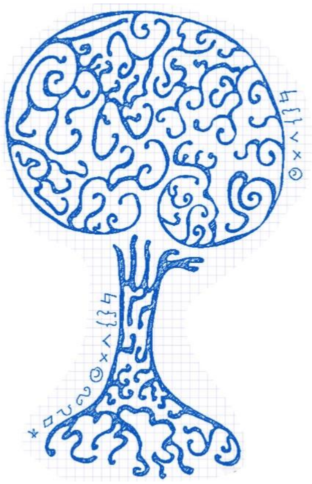
Overall, our results suggest that students exhibiting lower reading abilities in Grade 1 are the ones who show bigger learning improvements during the first two years of primary school. Lower reading ability in Grade 1, the period of time when students are starting to receive formal instruction in reading, can be related to a variety of factors, such as students' inequality in the amount of informal reading instruction received during the pre-primary school years, different experience with printed words, underlying reading disorders etc. Previous literature has suggested that typically developing students who learn how to read later than their peers can catch up with students who have more experience in reading by the age of 11 (Suggate et al., 2013). Our results provide support to this finding in two ways. First, by indicating that both in tasks assessing orthographic lexical processing and in the standardized measures of reading accuracy and reading time, lower performing students in Grade 1 were the ones improving more during the first years of primary school. Second, by showing that children's standardized reading performance exhibited lower interindividual variability in Grade 3, compared to Grade 1 students.

To sum up, the findings of the present chapter laid the groundwork in order to investigate in the next chapters whether students' metacognitive ability in first years of primary school aligns with the pattern of improvement in students' task performance both in linguistic and non-linguistic contexts. Moreover, they set the base to investigate whether metacognitive abilities in reading-related tasks significantly contribute not only to students' future reading performance, but also to students' learning improvements over time.

Chapter II Summary

- The present Chapter explored associations between type-1 performance in tasks related to orthographic lexical processing (VA span, lexical decision) and standardized reading ability tests (reading accuracy and time) in order to set the bases to subsequently explore in the following chapters the role of metacognition in these tasks.
- Students' performance in the lexical decision task, reflecting their orthographic knowledge, significantly correlated with their standardized reading ability both in Grade 1 and Grade 3. This result is in line with previous literature suggesting that orthographic knowledge correlates with single-item reading and word recognition (Katz et al., 2012).
- Students' performance in the VA span task, a prerequisite of orthographic lexical processing, significantly or marginally significantly correlated with their standardized reading ability in Grade 1, but not in Grade 3, indicating that the contribution of VA span in single-item reading is stronger in the first year of primary school, when children first receive formal instruction to reading (Bosse & Valdois, 2009).
- As expected, no association was found between the non-linguistic task (emotion recognition) and the reading variables.
- A significant improvement in all reading-related variables was shown between Grade 1 and Grade 3, following findings of previous studies suggesting that markers of orthographic lexical processing and single-item reading show an accelerated improvement across the first 3 years of primary school (Bosse & Valdois, 2009; Popa, 2020).
- Early performance in tasks related to orthographic lexical processing (VA span, lexical decision) negatively predicted longitudinal improvements within each task and in standardized reading tasks. Moreover, inter-individual variability in reading measures decreased over time. These findings indicate that less skilled readers in Grade 1 are the ones improving more over time to catch up with their more experienced peers (Suggate et al., 2013).

3. Chapter III: Developmental trajectories of metacognition in the early stages of primary school: Insights on the domain generality/specificity debate.



3.1. Introduction

The present chapter will investigate: a) how metacognitive efficiency develops over time and b) whether the metacognitive system is supported by domain general or domain specific mechanisms across the different tasks and critically whether there are any developmental changes regarding the domain generality of metacognition in the early stages of primary school. Below I outline the critical background knowledge and the questions addressed in this Chapter.

3.1.1. Does metacognitive efficiency change significantly across time during the first years of primary school? If yes, are these changes fueled by students' task sensitivity in the beginning of primary school?

As presented in the General Introduction, evidence for early metacognitive processing during development has been found in infants (Goupil & Kouider, 2016), while other studies have reported that the ability to use explicit confidence judgments to efficiently track task performance develops around the age of 5 (Destan et al., 2014; Rohwer et al., 2012). Children have been considered to start using this information in order to control their performance at the age of 6 (Destan et al., 2014; Rohwer et al., 2012), which coincides with the first year of schooling that has been suggested to bring a shift in students' cognitive abilities (Brod et al., 2017). Notwithstanding, understanding of metacognition's developmental course during the first years of primary school remains incomplete. Crucially, how experience in a certain task affects the developmental trajectories of metacognitive processing within a population has been scarcely studied.

Here, we aim to longitudinally examine the properties of metacognitive processing across different linguistic and non-linguistic tasks during the first years of primary school. To our knowledge, there are very few studies using a within-subject longitudinal design to assess the development of metacognition, and most of them involve studies on declarative metacognition (Lecce et al., 2015; Lockl & Schneider, 2007). Roebers and Spiess (2016) were among the first ones to assess type-1 and type-2 online performance of Grade 2 students in a spelling task across time (same population assessed in the beginning and the end of the school year) and suggested that there is a significant increase of students' metacognitive monitoring discimination across time, despite a significant decrease in overall confidence (Roebers & Spiess, 2017). In this study, metacognitive monitoring was measured as the difference between the mean confidence of correct and incorrect responses. However, this is a metacognitive index which does not avoid the confounding effects of confidence bias. The added value of our study is that: a) it uses an hierarchical, free of bias, Bayesian framework to estimate type-1 performance (d-prime) and type-2 performance (metacognitive sensitivity - meta-d' and metacognitive efficiency - meta-d'/d'), b) it tracks the development of metacognition across different linguistic and non-linguistic tasks during the first 3 years of primary school within the same subjects, longitudinally (Grade 1 to Grade 3).

Metacognitive efficiency is a recently developed metacognitive index (Fleming & Lau, 2014) and hence has not been extensively studied to assess participants across the life span. Palmer et al. (2014) were the first ones to study metacognitive efficiency in a visual perceptual and a memory task across time in age groups of healthy adults between 18-84 years old and reported a significant decrease in metacognition across time (Palmer et al., 2014). To our knowledge, there is only one study assessing metacognitive efficiency during development, which suggests that this index increases between childhood (8-9 years old) and adolescence (12-13 years old) and remains stable until late adolescence (16-17 years old) (Moses-Payne et al., 2021). The current study is to our knowledge the first study to assess metacognitive efficiency earlier on during childhood and by using a within-subject longitudinal design.

Assessing metacognition longitudinally entails the challenge that metacognitive ability is bound to task performance. When individuals become more competent in type-1 sensitivity in the task, it is expected that type-2 sensitivity will also increase (trivially because the type-1 signal available for the metacognitive monitoring process is higher) and hence, it is hard to disentangle the effect of experience in a task from the effect of age on metacognition (Roebers et al., 2019). In the present study we aimed to investigate how metacognitive efficiency, an index of metacognition which controls for type-1 sensitivity confounds, develops during the first years of primary school.

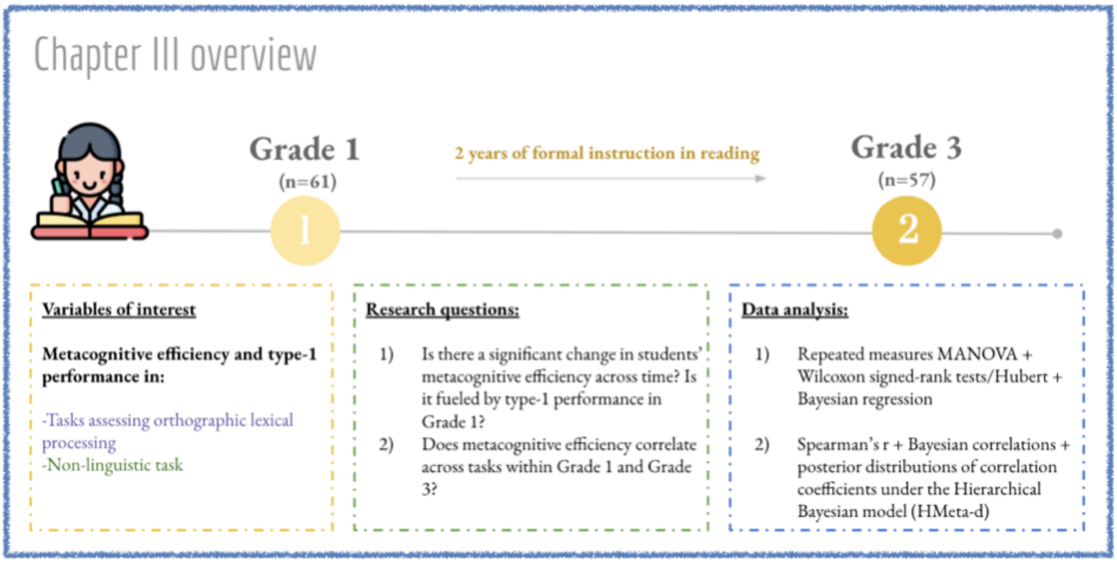
Next, we sought to examine whether changes in metacognitive efficiency across time are bound to early task sensitivity. Roebers' research group (2016), found that type-1 performance in a spelling task in Grade 2 students (8-9 years old) predicted not only participants' future performance in the task 8 months later, but also future metacognitive performance, suggesting that type-1 performance can be a driving force in the development of metacognition. If early task sensitivity can predict not only future metacognitive performance, but also developmental changes in metacognition between distinct time points, we expect that early task sensitivity will be a significant predictor of longitudinal changes in metacognitive efficiency. In later chapters of this thesis, we will also explore whether the relationship between type-1 and type-2 performance development across time is reciprocal, or whether one of the two is a stronger predictor of future learning.

3.1.2. Is metacognitive efficiency supported by domain specific or domain general mechanisms in the first years of primary school?

As mentioned in the General Introduction, the few existing developmental studies addressing the question of domain generality/specificity suggest that metacognitive processing is domain-specific in early childhood and becomes domain-general later during development (Geurten et al., 2018; Lyons & Ghetti, 2010; Vo et al., 2014). Specifically, Vo et al. (2014) showed that metacognitive sensitivity in 5-8 year olds was unrelated between numerical and emotion discrimination tasks (Vo et al., 2014). Geurten et al (2018) assessed metacognition in arithmetic and memory tasks in a group of children aged between 8 and 13. They observed the presence of a gradual shift towards a domain-general metacognitive system after the age of 10, with larger correlations between metacognitive indexes in the age group of 12-13 years old compared to the group of 10-11 years old.

No correlation was detected in younger age groups (Geurten et al., 2018). Following this study, Bellon et al. (2019) recruited two groups of children aged 7-8 and 8-9 years old and examined metacognitive processing across arithmetic and spelling tasks. The results suggested that already from the age of 8, metacognitive ability was correlated between the two domains, and this was predictive of the academic achievement in these domains (Bellon et al., 2019; Geurten et al., 2018). However, these studies are limited in the usage of metacognitive indexes that are sensitive to one's confidence biases or that may be confounded by the level of type-1 task performance (see Fleming & Lau, 2014).

The present study will address these issues by using state-of-the-art measures of metacognitive function (see Fleming & Lau, 2014), also in combination with Bayesian correlational analysis, which provide evidence in respect to the null hypothesis. To our knowledge, this is the first study to longitudinally assess this question in development. We expect that metacognitive efficiency will not correlate across tasks in the first wave of the study when students are 6-7 years old, as previous literature has suggested that domain-specific mechanisms support metacognition early in development (Geurten et al., 2018; Lyons & Ghetti, 2010; Vo et al., 2014). Next, we aim to explore if domain domain specific/general mechanisms support metacognition in the second wave of the study (age of students: 8-9 y.o.) to shed light on contradictory evidence of previous literature (see above) on whether or not there is a shift to domain general mechanisms already from this age.



Variables of interest

Metacognitive efficiency and type-1 performance in:

-Tasks assessing orthographic lexical processing
-Non-linguistic task

Research questions:

1) Is there a significant change in students' metacognitive efficiency across time? Is it fueled by type-1 performance in Grade 1?
2) Does metacognitive efficiency correlate across tasks within Grade 1 and Grade 3?

Data analysis:

1) Repeated measures MANOVA + Wilcoxon signed-rank tests/Hubert + Bayesian regression
2) Spearman's r + Bayesian correlations + posterior distributions of correlation coefficients under the Hierarchical Bayesian model (HMeta-d)

Figure 13 Schematic representation of the research design used in Chapter III.

3.2. Methods

3.2.1. Participants

As mentioned in the Chapter of Methodological Considerations, sixty-one children, native Spanish speakers, aged between 6 and 7 years (mean age (±SD): 6.59 ± 0.29, 28 girls), were assessed in the middle of Grade 1 and included in the data analysis of the first branch of the present longitudinal study. The second assessment point of this study occurred 2 years later, in Grade 3, when the sample consisted of fifty-seven children (mean age (±SD): 8.48 ± 0.34, 25 girls). The reduction of the sample was due to outliers identified in the first time point of the study, that were not re-assessed in Grade 3 or participants' change of school. A detailed description of participants' characteristics and process of recruitment can be found in Chapter Section 1.3.2.1.

3.2.2. Materials and Methods

Participants performed two linguistic tasks related to orthographic lexical processing; a VA span task, assessing the homonymous prerequisite skill of orthographic lexical processing and a lexical decision task, assessing children's orthographic lexical knowledge, and one non-linguistic, emotion recognition task. Moreover, a task assessing non-verbal reasoning abilities was administered to be used as a control task. The details of all tasks administered and the assessment process can be found in Chapter Section 1.3.2.3.

3.2.3. Data analysis

As mentioned in Chapter Section 1.3.2.4., participants' type-1 performance (d-prime) and type-2 performance (metacognitive sensitivity - meta-d' and metacognitive efficiency - meta-d'/d' or Mratio), were estimated under the recently developed hierarchical Bayesian model (Fleming, 2017). Mratio was estimated both in a single-subject level and in a group level in the three experimental tasks (VA span, lexical decision, emotion recognition). Group level estimates constrain the impact of single-subject estimates with high uncertainty on the group and have been considered particularly useful in studies including few trials like ours, and in examining whether domain general/specific mechanisms are recruited in tasks, by calculating the covariance between the estimates. However, group level estimates can only be used for Bayesian hierarchical modelling.

Here, in order to complement the analysis with post-fit frequentist analysis, we also calculated the single-subject parameter estimates of metacognitive efficiency.

First, we aimed at identifying possible differences in participants' type-1 (d-prime) and type-2 (meta-d', metacognitive bias and Mratio) individual performance across time (from Grade 1 to Grade 3). To this end, one-way repeated measures MANOVAS were applied for each variable (d-prime, meta-d', metacognitive bias, Mratio) using the SPSS software, Version 28.0. When a main effect of time was discovered, Bonferroni corrected post hoc tests were applied. In case of violation of sphericity assumptions, Greenhouse-Geisser correction was used to correct p-values. In case of violation of normality assumption, log transformation of the data was applied, and when normality was still violated in some conditions, post-hoc analyses were repeated on those using non-parametric Wilcoxon signed-rank tests to confirm the pattern of results. Subsequently, Spearman's r correlations in type-1 and type-2 performance estimates were applied to investigate possible correlations of these variables within and across the time points of the study. For each correlation analysis, False Discovery Rate (FDR) was used for multiple comparison correction and participants' chronological age and intellectual ability were used as covariates. Bayesian factors (BF_{10}) were also calculated for each pair of analyses in order to provide further evidence towards or against the null hypothesis (Andraszewicz et al., 2015).

Next, in order to assess differences in participants' metacognitive efficiency across developmental time points under the Bayesian hierarchical framework, we calculated the difference between the posterior distributions of the group-level estimates of metacognitive efficiency in each time point of the study within a task, estimated using the Hmeta-d toolbox function: fit_meta_d_mcmc_group.m. Significant differences across time points are indicated when the 95% highest density intervals (HDI) of the difference of the posterior distributions do not overlap with zero.

After assessing how type-2 metacognitive efficiency changes across time, we investigated whether these changes between Grade 1 and 3 are linked to participants' early type-1 performance in Grade 1 by using linear and Bayesian regression analyses. For the linear regressions, Huber robust regression was applied, which accounts for outliers. For the Bayesian regressions, a default prior of 0.354, as implemented in JASP software (van Doorn et al., 2020) was used and the Bayesian inclusion factor ($BF_{inclusion}$) was estimated for every predictor in the model. $BF_{inclusion}$ is calculated by dividing the prior odds of a model including a predictor of interest by the posterior odds (i.e., BF_{10}) excluding this predictor. When $BF_{inclusion} > 1$, it indicates that the model was improved by the addition of this specific predictor.

Subsequently, in order to examine whether type-2 metacognitive efficiency of the participants correlated across tasks within each time point we applied Spearman's r correlations in the single-subject estimates of participants' metacognition. In order to verify these correlations using group-level estimates of type-2 metacognitive efficiency, we applied the HMeta-d toolbox function: fit_meta_d_mcmc_groupCorr.m to calculate the 95% highest density intervals (HDIs) on the posterior distributions of the correlations coefficients (for details see Fleming, 2017). Significance is indicated when the posterior distributions do not overlap with zero.

Finally, in order to investigate whether the strength and direction of the abovementioned correlations between each pair of tasks changes across the two time points of the study, we compared the Spearman's r correlation coefficients, using the on-line R package tool "cocor-comparing correlations" (Diedenhofen & Musch, 2015). This tool compares the correlation coefficients of dependent groups (here same group tested in different time points), using the following tests: Pearson and Filon's z test (Pearson & Filon, 1897), Dunn and Clark's z test (Dunn & Clark, 1969), Steiger's z test (Steiger, 1980), Raghunathan, Rosenthal, and Rubin's z test (Raghunathan et al., 1996), Silver, Hittner, and May's z test (Silver et al., 2004), Zou's confidence interval test (Zou, 2007). In order to be able to introduce in the "cocor" tool the Spearman's correlation coefficients between single-subject estimates of metacognitive efficiency across tasks and time points, we had to repeat correlations excluding all outliers from each task and time point to equalize the sample size across all correlations (n=37).

3.3. Results

3.3.1. Descriptives

Table 9 summarises the descriptive analysis for the measures used to assess type-1 performance (d-prime) and type-2 performance (meta-d', metacognitive bias, Mratio) in each task.

Table 9 Descriptive statistics (Mean Score (SD), Range, Skewness) for the different measures of participants' type-1 and type-2 performance in the experimental tasks.

	Linguistic tasks related to orthographic lexical processing				***Non-linguistic task***	
	VA Span		Lexical Decision		Emotion Recognition	
	Grade 1	Grade 3	Grade 1	Grade 3	Grade 1	Grade 3
Type-1 task sensitivity (d-prime)						
Mean (SD)	0.84 (0.29)	1.28 (0.25)	1.26 (0.47)	1.68 (0.43)	1.21 (0.18)	1.25 (0.21)
Range	0.29-1.45	0.83-1.86	0.30-2.37	0.86-3.07	0.94-2.05	0.92-1.89
Skewness	0.133	0.469	0.282	1.081	1.991	0.738
Type-2 Metacognitive Bias (overall mean confidence 1:low-2:high)						
Mean (SD)	1.87 (0.14)	1.74 (0.09)	1.81(0.18)	1.75 (0.08)	1.85 (0.15)	1.71 (0.15)
Range	1.30-2.00	1.55-1.99	1.15-2.00	1.55-1.88	1.27-2.00	1.13-1.89
Skewness	-2.429	0.083	-1.949	-0.552	-1.747	-1.373
Type-2 Metacognitive Sensitivity (meta-d')						
Mean (SD)	0.99 (0.55)	0.72 (0.75)	1.77 (0.94)	0.90 (0.74)	1.69 (0.76)	0.51 (0.75)
Range	-0.49-2.36	-0.84-2.58	0.14-3.82	-0.72-2.36	0.12-3.49	-0.87-2.83
Skewness	0.096	0.110	0.032	0.046	0.247	0.825
Type-2 Metacognitive Efficiency (Mratio)						
Mean (SD)	1.31 (0.87)	0.57 (0.61)	1.65 (1.48)	0.53 (0.44)	1.44 (0.70)	0.41 (0.62)
Range	-0.73-4.10	-0.85-2.28	0.12-9.32	-0.55-1.46	0.09-3.31	-0.78-2.43
Skewness	0.850	0.175	3.281	-0.063	0.433	0.808

3.3.2. Comparing participants' type-1 and type-2 performance across time (Grades 1 to 3) within each experimental task

Type-1 task sensitivity (d-prime)

This analysis was performed in the context of Chapter II, hence the results are only briefly mentioned in the present chapter for purposes of clarity (for further details see Chapter Section 2.3.3.). Type-1 task sensitivity (d-prime) was found to increase from Grade1 to Grade 3 within the VA span and the lexical decision task (all ps < 0.001), but showed no difference within the emotion recognition task ($p = 0.780$, see Figure 14a). Next, intersubject correlations showed that type-1 task sensitivity did not correlate across time points in any of the tasks (all ps > 0.05). A marginal positive correlation was indicated in the lexical decision task ($r = 0.412$, $p = 0.067$).

Type-2 metacognitive bias

Participants' type-2 metacognitive bias, measured as the overall mean confidence of participants, was compared across the three experimental tasks and across time. Some variables were not normally distributed, but remained as such even after log transformation was applied in the data, so it was decided to perform a parametric one-way MANOVA in the non transformed data, followed by non-parametric Wilcoxon signed-rank pair to pair tests.

A one-way repeated-measures MANOVA with time (Grade 1, Grade 3) as a within-subject factor on participants' type-1 task sensitivity in the three experimental tasks (VA span, lexical decision, emotion recognition), revealed a large main effect of the time point of the study ($F(3,34) = 11.845$, $p < 0.001$, $\eta_p^2 = 0.511$). In order to confirm that the time effect applied to all the tasks, univariate post-hoc analysis was performed, which revealed that metacognitive bias showed a significant decrease from Grade1 to Grade 3 in the VA span task and the emotion recognition task (all ps < 0.001), but not in the lexical decision task ($p = 0.280$, see Figure 14b). The post hoc analysis was repeated using pair-to-pair non-parametric Wilcoxon signed-rank tests and the same pattern of results was revealed.

Finally, intersubject correlations showed that type-2 metacognitive bias was highly correlated across tasks in Grade 1 (all ps < 0.001). In Grade 3, metacognitive bias significantly or marginally significantly correlated across all pair of tasks (VA span-lexical decision: $r = 0.466$, $p = 0.003$; VA span-emotion recognition: $r = 0.335$, $p = 0.052$; lexical decision-emotion recognition: $r = 0.466$, $p = 0.003$). However, no across time correlations were indicated in any pair of tasks (all ps > 0.35).

Type-2 metacognitive sensitivity (meta-d')

We next compared participants' type-2 metacognitive sensitivity across the three experimental tasks and across time. One variable was not normally distributed, but remained as such after log transformation was applied in the data, so it was decided to perform a parametric one-way MANOVA in the non transformed data, followed by non-parametric Wilcoxon signed-rank pair-to-pair tests.

A one-way repeated-measures MANOVA with time (Grade 1, Grade 3) as a within-subject factor on participants' type-1 task sensitivity in the three experimental tasks (VA span, lexical decision, emotion recognition) revealed a large main effect of the time point of the study ($F(3,34) = 13.882$, $p < 0.001$, $\eta_p^2 = 0.551$). In order to confirm that the time effect applied to all the tasks, univariate post-hoc analysis was performed, which revealed that metacognitive sensitivity showed a significant decrease from Grade1 to Grade 3 in the lexical decision and the emotion recognition task (all ps < 0.001), but marginally significant in the VA span task ($p = 0.074$, see Figure 14c). The post hoc analysis was repeated using pair-to-pair non-parametric Wilcoxon signed-rank tests and the same pattern of results was reproduced. Finally, intersubject correlations showed that type-2 metacognitive sensitivity did not correlate significantly across time points in any of the tasks (all ps < 0.05).

Type-2 metacognitive efficiency (meta-d'/d' or Mratio)

We next compared participants' type-2 metacognitive efficiency across the three experimental tasks and across time. Some variables were not normally distributed, but remained as such even after log transformation was applied in the data, so it was decided to perform a parametric one-way MANOVA in the non transformed data, followed by non-parametric Wilcoxon signed-rank pair to pair tests.

A one-way repeated-measures MANOVA with time (Grade 1, Grade 3) as a within-subject factor on participants' type-1 task sensitivity in the three experimental tasks (VA span, lexical decision, emotion recognition) revealed a large main effect of the time point of the study ($F(3,34) = 19.620$, $p < 0.001$, $\eta_p^2 = 0.634$). In order to confirm that the time effect applied to all the tasks, univariate post-hoc analysis was performed, which revealed that metacognitive efficiency showed a significant decrease from Grade1 to Grade 3 in all tasks (all ps < 0.001, see Figure 14d). The post hoc analysis was repeated using pair-to-pair non-parametric Wilcoxon signed-rank tests and the same pattern of results was revealed.

Next, the difference between the posterior distributions of the group-level estimates of metacognition was calculated between time points for each experimental task. The same pattern of results was revealed, with all HDIs of the difference not overlapping 0, indicating significant decrease in metacognitive efficiency between time points (VA Span (Grade 3 vs Grade 1): 95% HDI

= [-1.50, -0.48]; , Lexical Decision (Grade 3 vs Grade 1): 95% HDI = [-1.64,-0.77]; Emotion Recognition (Grade 3 vs Grade 1): 95% HDI = [-3.09, -1.30]).

Finally, intersubject correlations showed that type-2 metacognitive efficiency did not correlate significantly across time points within tasks (all ps < 0.05). Correlation was marginally significant in the VA span task after FDR correction (r = 0.377, p = 0.080). When evaluating the covariance of participants' group metacognitive efficiency across time points in each task within the hierarchical model of the Bayesian framework, no substantial covariance was suggested in any of the tasks. This was shown by the 95% HDIs on the posterior distributions of the correlation coefficients overlapping zero in all cases (VA Span (Grade 3 vs Grade 1): 95% HDI = [-0.79, 0.99]; Lexical Decision (Grade 3 vs Grade 1): 95% HDI = [-0.54, 0.91], Emotion Recognition (Grade 3 vs Grade 1): 95% HDI = [-0.99, 0.83]).

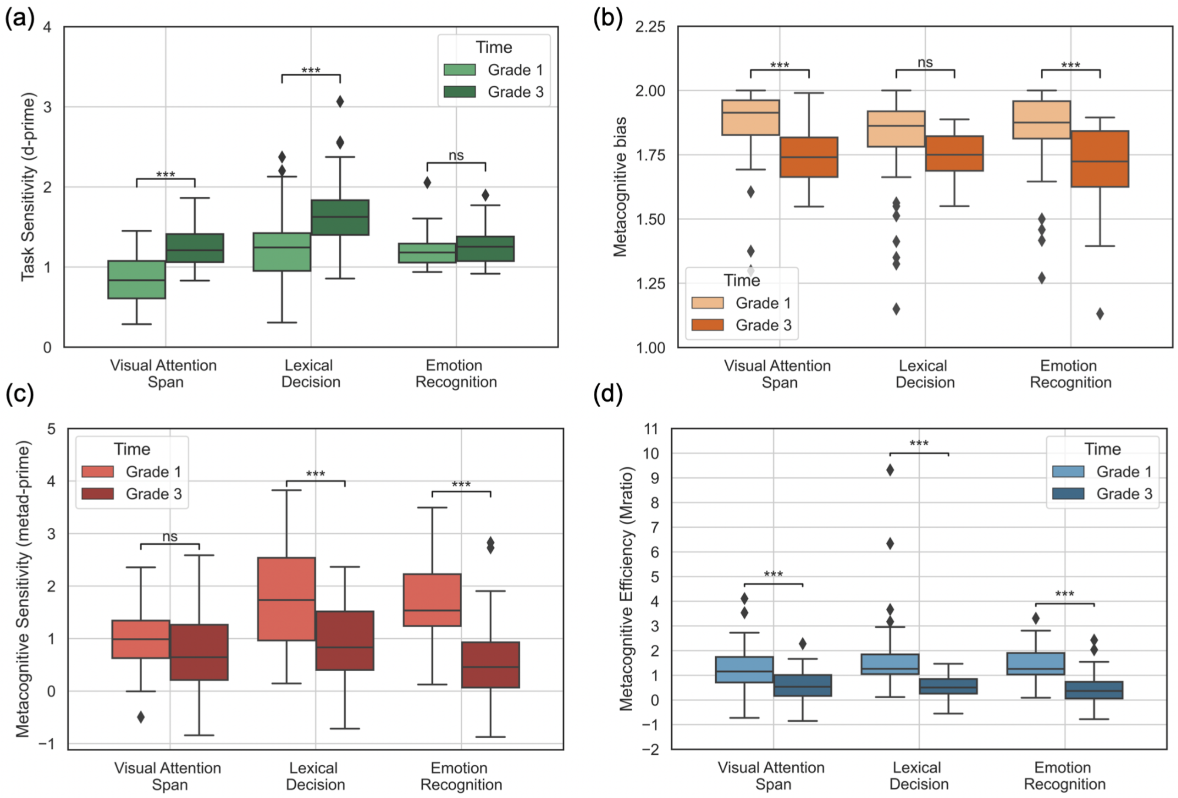


Figure 14 (a) Task sensitivity (d-prime), (b) metacognitive bias, (c) metacognitive sensitivity (meta-d') and (d) metacognitive efficiency (single-subject Mratio) in Grade 1 and Grade 3 of the longitudinal study within the different experimental tasks. Significance is indicated for the effect of the time point of the study (*p < 0.05, **p < 0.01, ***p < 0.001, Bonferroni correction).

3.3.3. Investigating the relationship between early task sensitivity and developmental changes in participants' metacognitive efficiency across time

We investigated whether task sensitivity (d-prime) of students in Grade 1 predicts longitudinal changes on their metacognitive efficiency (Mratio) during a period of 2 years across tasks. We conducted regression analyses to predict longitudinal changes in metacognitive efficiency (i.e., learning effects), measured as the difference in the Mratio between Grade 1 and Grade 3 for each task, using d-prime, age and IQ at Grade 1 as predictors. Analyses were performed using a robust to outliers regression model (Hubert regression) and Bayesian regression models, which indicated d-prime in Grade 1 as a significant predictor of Mratio longitudinal changes in the lexical decision and the emotion recognition task (all ps < 0.05, $BF_{10} > 1$, see Table 10). This result indicates that students with lower d-prime in Grade 1 in these experimental tasks, were the ones who showed the bigger reduction in metacognitive efficiency across time, while students with higher d-prime in Grade 1, showed smaller or no reduction of their metacognitive efficiency across time (see Figure 15). It is worth-mentioning here that in Chapter II, we showed that students' with lower d-prime in Grade 1, were the ones who improved more in their type-1 performance across time.

3.3.4. Assessing domain general contributions to metacognitive efficiency in Grade 1 and Grade 3

Spearman's correlations coefficients between single-subject estimates of metacognitive efficiency (Mratio) across the different tasks within Grade 1 and Grade 3 are presented in Figure 16. The goal was to examine whether metacognitive efficiency is supported by domain-general mechanisms. In Grade 1, a strong positive correlation was observed between metacognitive efficiency on the orthographic lexical decision task and the emotion recognition task ($r = 0.525$, $p < 0.001$). Bayes factor provides very strong evidence in favour of this hypothesis ($BF_{10} = 44.143$), suggesting the use of a common metacognitive mechanism in these tasks. A significant correlation was also noted between metacognitive efficiency on the VA span task and the orthographic lexical decision task ($r = 0.330$, $p = 0.033$), with the Bayes factor providing only anecdotal evidence towards the alternative hypothesis ($BF_{10} > 1$). In Grade 3, no significant correlation was observed between any pair of tasks (all ps > 0.60). In all cases, Bayes factor provided moderate evidence in favour of the null hypothesis (all $BF_{10} < 0.33$), suggesting the use of domain specific mechanisms supporting metacognition in this age.

Table 10 Regression analysis (Hubert regression) of participants' longitudinal changes in Mratio between the two time points of the longitudinal study in a) the VA Span task (n=47), b) the lexical decision task(n=48), c) the emotion recognition task (n=49) with task sensitivity, age and non-verbal IQ at Grade 1 as predictors. $BF_{inclusion}$ factor represents the change from prior to posterior probabilities of a model when a predictor is added in the equation ($BF_{inclusion} > 1$ indicates that the predictor improves the model).

(a)

	Mratio changes in VA Span task (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
VA Span d-prime (Grade 1)	0.615	1.414	0.163	0.751	1.095
Age	0.244	0.612	0.539	0.399	1.000
Non-verbal IQ	0.043	1.152	0.251	0.395	1.000

(b)

	Mratio changes in Lexical Decision task (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Lexical Decision d-prime (Grade 1)	0.648	2.161	0.040	5.291	7.778
Age	0.945	1.981	0.052	1.472	1.000
Non-verbal IQ	0.072	1.635	0.110	1.495	1.000

(c)

	Mratio changes in Emotion Recognition task (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Emotion Recognition d-prime (Grade 1)	1.727	2.587	0.013	0.690	1.654
Age	0.351	0.818	0.417	0.462	1.000
Non-verbal IQ	0.025	0.685	0.499	0.333	1.000

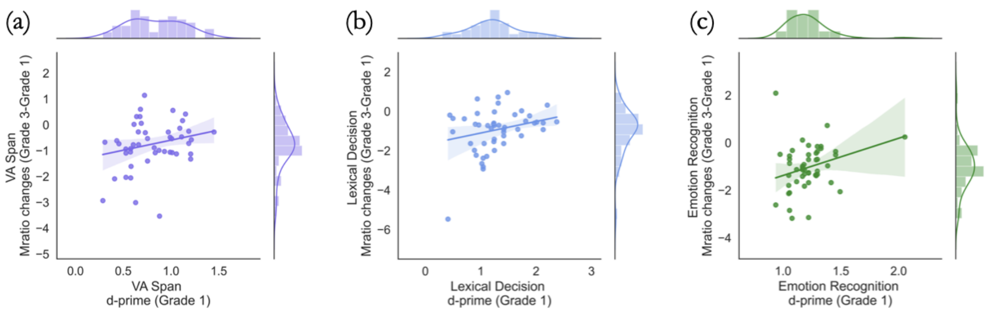


Figure 15 Linear relationship between longitudinal changes in Mratio across time and task sensitivity in Grade 1 (a: VA span task, b: lexical decision task, c: emotion recognition task).

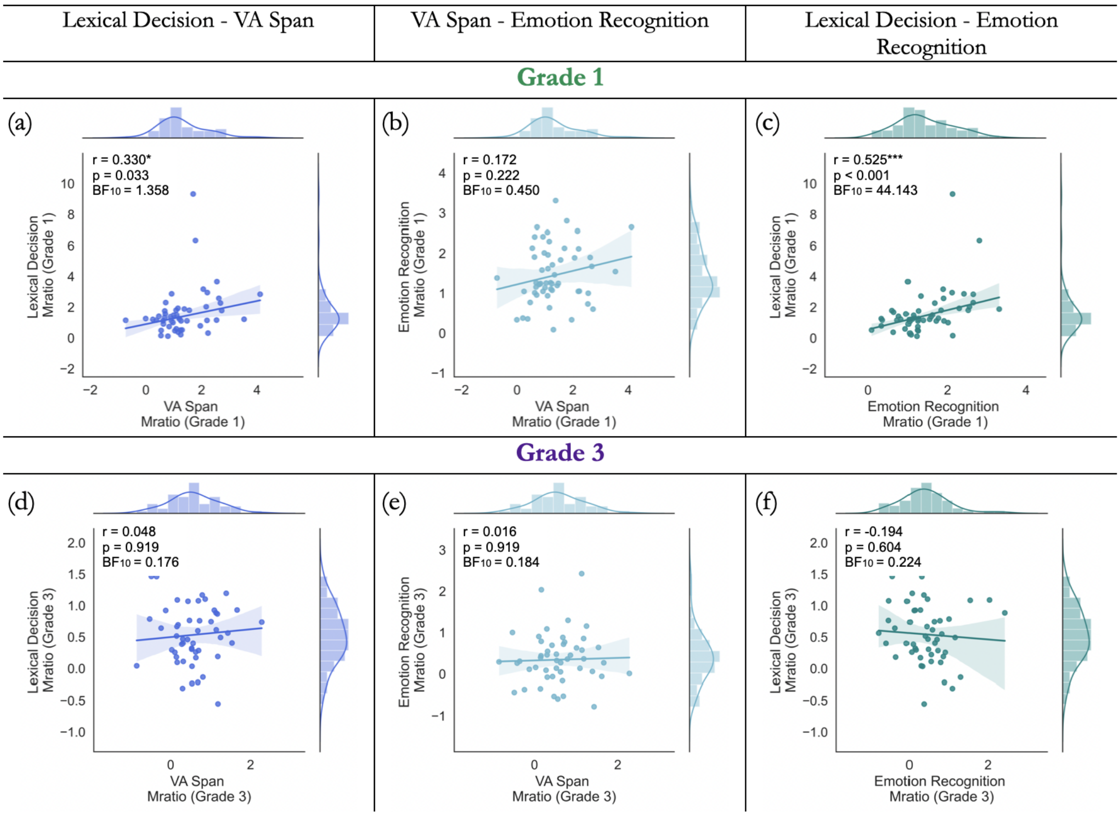


Figure 16 Spearman correlations among type-2 metacognitive efficiency (Mratio) in experimental tasks (*p < 0.05, **p < 0.01, ***p < 0.001, FDR corrected) in Grade 1 (a-c) and Grade 3 (d-f). Correlations were controlled for participant's age and intellectual ability (non-verbal IQ).

Next, the covariance of participants' group estimates of metacognitive efficiency across tasks was evaluated within the hierarchical model of the Bayesian framework. In Grade 1, substantial covariance was suggested only between the lexical decision task and the emotion recognition task, shown by 95% HDIs on the posterior distributions of the correlation coefficients which do not overlap zero, hence indicating a significant correlation (VA Span-Lexical Decision: $\rho=0.491$, 95% HDI=[-0.38, 0,99]; VA Span-Emotion Recognition: $\rho=-0.063$, 95% HDI = [-0.99, 0.74], Lexical Decision-Emotion Recognition: $\rho=0.842$, 95% HDI=[0.57, 0.99], see Figure 17). In Grade 3, no substantial covariance was suggested in any pair of tasks, shown by 95% HDIs on the posterior distributions of the correlation coefficients which overlapping zero in all cases, hence indicating as well non-significant correlations between tasks (VA Span-Lexical Decision: 95% HDI = [-0.60, 0.98]; VA Span-Emotion Recognition: 95% HDI = [-0.80, 0.95], Lexical Decision-Emotion Recognition: 95% HDI = [-0.80, 0.90], see Figure 17).

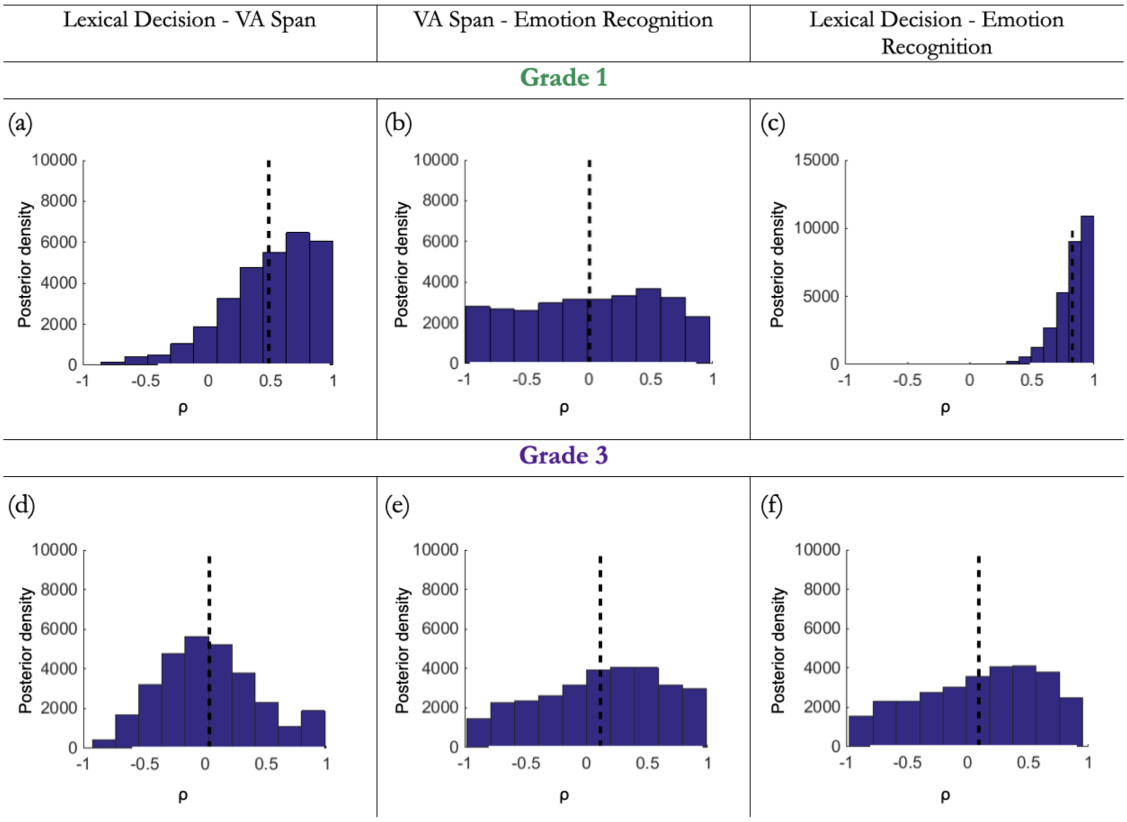


Figure 17 Posterior distributions over ρ for each correlation pair across the experimental tasks determining covariance between metacognitive efficiency across tasks in Grade 1 (a-c) and Grade 3 (d-f). The white horizontal bar indicates the 95% of high density intervals (HDIs). The black dotted line indicates the ground-truth correlation between type-2 metacognitive efficiencies. In cases where 95% HDIs on posterior correlation coefficients do not overlap zero, a substantial covariance in metacognitive efficiency across domains is suggested.

Moreover, we compared the magnitude of the correlations in each pair of tasks between the two time points of the study, using the R-package tool "cocor-comparing correlations" (Diedenhofen & Musch, 2015), after excluding all outliers to equalize the number of participants across tasks and time points ($n=37$). The magnitude of the correlation between metacognitive efficiency in the VA span and the lexical decision task was compared between Grade 1 and Grade 3 within the same group of children. All test results indicated that there was no significant difference between the two correlations in Grade 1 and Grade 3 and that the null hypothesis should be retained (all ps < 0.05). The same pattern of results was shown when comparing the correlation of metacognitive efficiency on VA span and the emotion recognition task between the two time points of the longitudinal study.

Finally, the magnitude of the correlation between metacognitive efficiency in the lexical decision and the emotion recognition task was compared between Grade 1 and Grade 3 and produced a significant difference by all tests used in “cocor” tool (all $ps > 0.05$), indicating a significant decrease in the magnitude of this correlation by Grade 3.

3.4. Discussion

The main goals of this chapter were twofold. First, we wanted to explore whether students' metacognitive efficiency significantly changes over the first years of primary school, while assessing the effect of early task performance on these changes. Second, we sought to determine whether metacognitive ability at this early stage of neurocognitive development is mediated by a domain-general or a domain-specific system, and critically, whether there is any developmental effect on the extent of the domain generality of the metacognitive system. Below, we briefly report the main results that will be discussed.

First, type-2 metacognitive efficiency significantly decreased over the first 2 years of primary school across domains. Critically, this decrease was noted even for tasks in which type-1 task sensitivity (d-prime) didn't change (i.e., emotion recognition task) or overall confidence and type-2 metacognitive sensitivity didn't show a significant decrease (i.e., lexical decision task) over time. Interestingly, early task sensitivity was a significant predictor of the decrease in metacognitive efficiency in the lexical decision and the emotion recognition task, but not in the VA span task, with participants' exhibiting lower type-1 task sensitivity (d-prime) in Grade 1, showing the biggest reduction in metacognitive efficiency across time and students with higher task sensitivity showing either smaller reduction or improvement of their metacognitive efficiency across time. Second, the levels of metacognitive efficiency on the lexical decision and the emotion recognition tasks were strongly associated both in an individual and group level analyses only in Grade 1. There was no other evidence of domain general metacognitive mechanisms supporting metacognitive efficiency across any pair of tasks in the first 2 years of primary school, notwithstanding domain generality of overall confidence shown within time points across all tasks. The magnitude of the correlations in metacognitive efficiency between the different pairs of tasks across time only changed in the case of the lexical decision and the emotion recognition task showing a significant decrease. This result was unexpected based on the existing evidence suggesting a gradual shift towards domain generality in metacognition (Bellon et al., 2019; Geurten et al., 2018; Vo et al., 2014).

Developmental changes of metacognitive efficiency in the first years of primary school

Metacognitive ability has been considered to develop over time and to present substantial advancements in its explicit form after the age of 5 (Destan et al., 2014; Rohwer et al., 2012). However, very few studies have tracked the progress of metacognition in the first years of primary school and existing studies differ in their use of metacognitive indexes. Destan et al. (2014) compared confidence ratings in correct vs incorrect trials and suggested that children after the age of 5 are more likely to rate with high confidence a correct than an incorrect response, but that age does not have an effect on this possibility in the first two years of primary school, even if children improve in the type-1 task performance. On the contrary, Roebers and Spiess (2017) using a longitudinal within-subject design found that metacognitive monitoring increases in students over the course of Grade 2 (Roebers & Spiess, 2017). However, both of these studies use metacognitive indexes that do not avoid the confounding effects of confidence bias and type-1 performance.

Here, we assessed children's type-1 and type-2 task performance using a hierarchical Bayesian framework, free of biases. We found that children became more competent in type-1 task performance (d-prime) in the linguistic tasks over the course of two years, even in the case that a staircase was used to initially adjust accuracy at 70%. Type-1 d-prime remained the same in the emotion recognition task. However, in Chapter Section 2.3.3., we showed that mean stimulus presentation time decreased in the same task across time, indicating that students also improved in this task.

Overall confidence decreased over time in the majority of the experimental tasks (except the lexical decision task), a result which comes in line with previous literature, suggesting that children's tendency to overestimate their confidence (Finn & Metcalfe, 2014) decreases between the ages of 7 to 10 years of age (Schneider, 2015). Interestingly, metacognitive efficiency robustly decreased over the first two years of primary school across all experimental tasks, even when type-1 performance or type-2 metacognitive sensitivity did not change significantly across time. This result comes in contrast to Roebers and Spiess (2017) who longitudinally assessed metacognitive monitoring, and suggested that it increases over time in the ages of 8 to 9 years old. One explanation for the apparent discrepancy in results is that different metacognitive indexes were used across the two studies. The robust decrease of metacognitive efficiency, a bias-free metacognitive index, across time may be related to students' previous experience on the task. Recently, Bang et al. (2019) suggested that meanwhile participants accumulated experience in performing a visual perceptual task, their type-1 sensitivity improved over time, while their metacognitive efficiency decreased over 7 sessions of testing on different days. This decrease in metacognitive efficiency was accompanied by a decrease in sensory noise. Sensory noise has been considered to affect type-1 decisions, with less sensory noise leading to higher task sensitivity. However, Bang et al. (2019) proposed that type-2 decisions are also affected by sensory noise, with higher sensory noise leading to more accurate metacognitive judgments. Notwithstanding, their findings suggested that in addition to the effect of sensory noise on metacognitive efficiency, there is another source of "metacognitive noise" affecting type-2 decisions together with the sensory noise. Metacognitive noise has been defined as "a type of noise that affects confidence judgments, but not participants' perceptual decisions" and has been attributed to several non-perceptual factors (e.g. confidence, trial/performance history, arousal etc., see: Shekhar & Rahnev, 2021; Bang et al., 2019). Hence, we suggest that the co-occurence of these processes (decrease in sensory noise, increase in metacognitive noise across time) may have led to the decrease of participants' metacognitive efficiency across tasks in our study. In other words, as students become more competent in type-1 performance and gain experience on a task, they may have less need of employing metacognitive resources to control their performance. We can understand this in real life settings, that when we excel in an activity, such as driving, we have less need to actively reflect on our performance. On this account, it is not the case that metacognitive performance reduces across development but simply that there may be less reliance on the metacognitive system as children gain experience at performing a task, which allows them to perform in a more automatic fashion. For instance, children during the very early stages of reading acquisition depend upon analytic reading or decoding of printed text, which requires more controlled attentional resources. When orthographic lexical processing is mastered, students rely more on sight-word reading, which is a rapid and automatic process. Sight word reading has been suggested to require minimal or non conscious effort upon familiar reading items (Hains, 1986; Samuels & Flor, 1997).

Of note, a difference between our study and Bang et al.'s findings is that mean confidence was here found to decrease between Grade 1 and Grade 3, while in Bang et al.'s study, mean confidence increased across the 7 different sessions alongside type-1 sensitivity. However, the observed decrease in children's confidence may well be attributed to developmental changes related to the natural decrease in wishful thinking and overconfidence tendency that has been suggested to occur between the ages of 7 to 10 years old (Schneider, 2015). Additional work is necessary to pinpoint the contribution of factors associated with metacognitive noise and confidence biases in early childhood for the development of metacognition. Also, further work is needed to make further determinations on whether the observed changes in type-1 and type-2 performance indexes are solely related to children's enhanced experience on the specific tasks, or are due to developmental changes that occur naturally. Longitudinal within-subject studies entail the issue that it is hard for one to disentangle the different factors that are driving an effect. On the contrary, using a cross-sectional study would introduce the problem of interindividual variability, but I suggest that including a control group in our research design of Grade 3 students who have no experience in the tasks would help us unravel the different factors underlying these effects.

A follow-up analysis indicated early type-1 task performance as a significant predictor of the decrease in metacognitive efficiency across time in the lexical decision and the emotion recognition task. In Chapter II, we provided evidence that early type-1 task performance is also a significant predictor of learning improvements in type-1 performance across time. Taken together, these results provide evidence for the hypothesis that domain specific knowledge can fuel changes both in type-1 and in type-2 performance (Schneider, 2015; Roebers' et al., 2016). The added value of the present study is that it assesses the relationship of early task performance with the quantifiable change in students' metacognition across time, instead of correlating those variables at different time points of a study, as in (Roebers' et al., 2016).

Domain general/specific mechanisms supporting metacognition in the early stages of primary school

A further goal of the present study was to investigate whether metacognitive ability in early childhood is supported by domain-general or domain-specific mechanisms. This issue has been scarcely studied in the field of cognitive development. Our data only revealed small evidence pointing to common underlying mechanisms supporting metacognition. In Grade 1, only in the lexical and emotion recognition tasks, participants' metacognitive efficiency, both in a single-subject and a group level, was highly correlated. At the same time point, we observed a weak correlation between single-subject estimates of metacognitive efficiency on the VA span task and the orthographic lexical decision task, which was not borne out by the analysis of group-level estimates under the hierarchical Bayesian framework. No correlation was found between the VA span and the emotion recognition task. In Grade 3, no association between any pairs of tasks was found.

Previous studies point to a gradual shift towards a domain general metacognitive system during childhood (Bellon et al., 2019; Geurten et al., 2018; Vo et al., 2014). First, Vo et al. (2014) suggested the existence of domain specific metacognitive mechanisms supporting numeric and emotional domains in the age of 5-8 (Vo et al., 2014). Geurten and colleagues later evaluated metacognition in different age groups in arithmetic and memory domains and suggested that the shift towards domain general mechanisms underlying metacognition is happening at the age of 10-11 (Geurten et al., 2018). A following study of Bellon et al (2019) found that correlations of metacognitive ability across arithmetic and spelling domains can already be detected from the age of 8-9 (Bellon et al., 2019). These studies are to our knowledge the only developmental studies studying cross-domain metacognition in different tasks using confidence judgments, but they are limited by the use of metacognitive indexes which do not control for the effect of metacognitive bias or the level of type-1 performance. Here, this issue was addressed by using group-level estimates of type-2 metacognitive efficiency (Mratio) under the Bayesian H-metad framework, which revealed little evidence for the existence of a common underlying mechanism of metacognition from the age of 6-7 only between one pair of tasks. However, no association was revealed between any pair of tasks in Grade 3 (8-9 year old children), contrary to Bellon et al. (2019) findings suggesting a shift to domain general mechanisms supporting metacognition already from the age of 8.

One explanation for the significant association noted between metacognitive performance in the lexical decision and the emotion recognition task in Grade 1, a pattern which was not repeated in Grade 3, may be based on the different cues used by the children in the two time points of the study in these two tasks. Strategic cue utilization and cue validity has been suggested to grow especially in younger children (before the age of 11) and to improve monitoring accuracy (Ackerman & Koriat, 2011; Roebers et al., 2019).

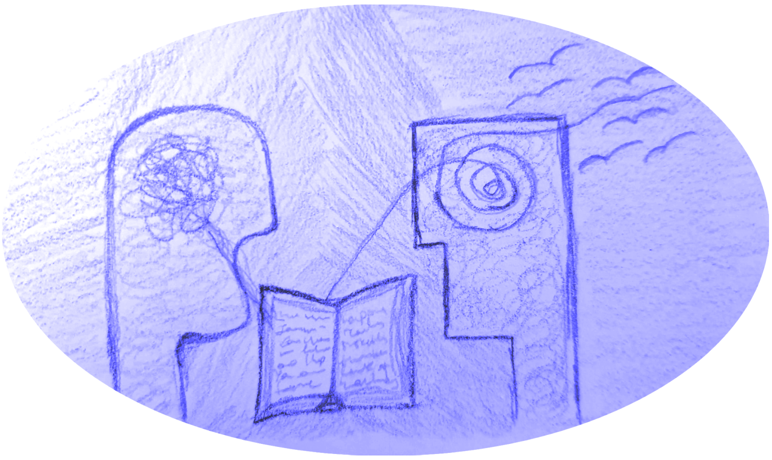
The lexical decision and the emotion recognition tasks are characterized by similar structure, both using a 2-alternative discrimination task design and a staircase to adjust type-1 performance. Specifically, the existence of a staircase procedure provided a variety of presentation duration timings of the stimuli during a task, which may have been similarly used in both tasks in Grade 1 as cues to inform childrens' confidence judgments. Hence, the existence of this strong correlation may be related to the fact that participants apply common heuristics in their metacognitive monitoring in these two tasks in Grade 1, rather than indicating the existence of a domain general mechanism supporting metacognition. Moreover, the absence of correlation in metacognitive indices Grade 3 may indicate that heuristics for metacognition vary across development based on participants' experience/knowledge on a certain domain or that childrens' repeated experience with tasks allow them to discover and use inherent task characteristics as cues to inform their confidence (Roebers et al., 2019).

Finally, another aspect of the study that needs to be considered is that participants' metacognition in the VA span task was assessed through a target detection ('Yes/No') task. Maniscalco and Lau (2011) have reported that in this type of tasks, metacognitive sensitivity (meta-d') for "no" responses is lower than for "yes" responses, as if the presence of the key target feature weights more the sensory representation than its absence (Maniscalco & Lau, 2011). Recently, it has been suggested that differences in task structure might hinder the detection of cross domain correlations (Ruby et al., 2017; Samaha & Postle, 2017). Mazancieux and colleagues (2019) showed cross-task correlations in metacognitive efficiency across four different tasks (i.e., semantic memory, episodic memory, executive function, visual perception), and all of the tasks with the same task structure (Mazancieux et al., 2020, Preprint). It would be relevant for future studies to re-examine the domain-generality issue during early neurocognitive development by using similar task structures across all cognitive domains.

Chapter III Summary

- Contrary to previous literature suggesting that metacognitive monitoring increases over time (Roebers & Spiess, 2017), we found that students' metacognition, indexed as metacognitive efficiency (Mratio), significantly reduced between Grade 1 and Grade 3 in al linguistic and non-linguistic tasks. We suggest that this reduction may be related to children's increased level of automaticity due to experience with these tasks, which allows them to rely less on their metacognitive system. This finding is in line with a recent research study suggesting that the decrease in type-1 sensory noise may be related to an increase in metacognitive noise (Shekhar & Rahnev, 2021).
- Students' with lower early task sensitivity in Grade 1 were the ones who showed the biggest reduction in their metacognitive efficiency, but also higher learning improvements in type-1 performance across time, both in linguistic and non-linguistic tasks. On the contrary, their higher performing peers in Grade 1, exhibited smaller or no reduction in their metacognitive efficiency across time. This finding may indicate that students who were performing worse in Grade 1, relied more on their metacognitive resources to catch up with their peers, and freed-up these resources over time, to devote them to type-1 performance.
- Little evidence of common mechanisms supporting metacognition in the lexical decision and the emotion recognition task was found in Grade 1. However, no other association was found in students' metacognitive efficiency in any other pair of tasks or time point of the study. These results are more in favour of domain-specific mechanisms supporting metacognition in the first years of primary school (Bellon et al., 2019; Geurten et al., 2018; Vo et al., 2014). The single association between the lexical decision and the emotion recognition task in Grade 1 may well be related to the identical structure of these two tasks. For instance, students in Grade 1 may have used the variety of presentation duration timings of the stimuli, as cues to inform their confidence ratings. Strategic cue utilization has been suggested to develop especially during the first years of primary school (Ackerman & Koriat, 2011; Roebers et al., 2019). Hence, in Grade 3, students may have used different heuristics in these two tasks during metacognitive monitoring.

4. Chapter IV: The role of metacognition in reading development



4.1. Introduction

The previous chapters provided evidence that students' type-1 performance in tasks related to orthographic lexical processing (and not in non-linguistic tasks) is linked to their reading ability as measured by standardized tests, mostly in the early stages of primary school (Grade 1, see Chapter II). Moreover, we showed that children exhibiting weaker reading abilities in the early stages of primary school are the ones who show more learning improvements across time in type-1 performance (Chapter II), but also bigger reduction in their type-2 metacognitive ability across domains (Chapter III).

In the present chapter we aim to assess:

a) the relationship between students' metacognitive ability in tasks related to orthographic lexical processing (VA span task, lexical decision task) and students' performance in standardized reading tests (reading accuracy and time) within and across time points of the study (Grade 1, 3).

b) the relationship between metacognitive efficiency in the VA span task, a precursor of orthographic lexical processing, and task sensitivity on the lexical decision task, reflecting orthographic knowledge within the two time points of the longitudinal study (Grades 1, 3).

c) the predictive value of type-2 metacognitive efficiency in reading-related experimental tasks in Grade 1 on student's performance improvements in type-1 performance in the experimental tasks and in standardized reading ability across time points (Grades 1-3).

The following research questions will be addressed:

4.1.1. Does metacognition contribute to reading performance and the development of the orthographic lexicon at earlier and later developmental stages during the primary school period?

In addition to reading, metacognitive ability has been considered fundamental for learning in other domains such as mathematics, memory and perception (Kuhn, 2000; Schoenfeld, 2016). Educational studies suggest that individuals with higher performance monitoring skills tend to be better learners (Metcalfe & Kornell, 2007; Rawson et al., 2011). However, to date, research in development is mainly based on self-report questionnaires and there is a lack of robust metrics of metacognition that can be comparable across tasks. Only recently, Bellon et al.'s (2019) study in early childhood (7-9 years of age) showed that metacognitive processing, in the context of spelling and arithmetic task performance, correlated with standardized tasks examining the level of performance in these domains (i.e., Spelling: standardized dictation task, see Moelands & Rymenans, 2003; Arithmetic: Tempo Test Arithmetic, see De Vos, 1992). Bellon et al. used a "metacognitive monitoring score", that measures the alignment between confidence judgments and accuracy (Bellon et al., 2019). Vo et al.'s (2014) examined metacognition in the numerical and emotion domain in the ages of 5-8, and found that metacognitive sensitivity on numerical judgments was positively correlated with math ability, but this was not the case with metacognitive sensitivity in the emotion domain (Vo et al., 2014). However, the abovementioned studies used metacognitive indexes which do not avoid confounding effects stemming from confidence bias (Bellon et al., 2019) or type-1 performance (Vo et al., 2014). In this study, we use free of bias metacognitive indexes, in order to measure more precisely the contribution of metacognition in the acquisition of orthographic knowledge that subtends fluent reading development.

Our first hypothesis is that, if metacognition is related to reading and its orthographic prerequisites, participants with higher metacognitive efficiency on the tasks indexing orthographic knowledge (i.e., VA span and orthographic lexical decision) will exhibit higher type-1 performance across these tasks and higher performance in the standardized reading tasks in both time points of the study (Grade 1 and Grade 3).

Recently, Filevich et al. suggested that the age of 5 to 6 is a critical age window in the development of metacognitive monitoring. Using tasks in which children had to recognize and report their knowledge certainty in the task, they showed that children's ability to correctly identify and explicitly report that they did not know is associated with key changes in cortical thickness in the medial orbitofrontal cortex (Filevich et al., 2020). Brod et al. (2017) have suggested that the first year of schooling brings a shift in children's cognitive abilities that may be critical for metacognition. Specifically, over this year, students showed great improvements in tasks requiring executive control functions, which are also linked to activity changes in parietal cortex regions associated with attention control. Brod et al. also used fMRI to show that these children also display increased activation of brain areas related to sustained attention (Brod et al., 2017). Our second hypothesis is that, if metacognitive monitoring contributes to the outlined "5-to-7 year shift" in cognitive abilities, we expect to see stronger associations between students' metacognitive efficiency supporting orthographic processing in the related tasks (VA span, lexical decision) and students' task performance in the same tasks and the standardized reading ability tasks in the first wave, comparing to the second wave, of the longitudinal study (children's age: 6-7 years old).

Finally, as mentioned in the Introduction, VA span has been indicated as a significant precursor of orthographic lexical processing and knowledge. Our third hypothesis is that if metacognition is involved in the process of the development of children's orthographic lexicon, students' metacognition in their VA span skills, will significantly associate with their orthographic knowledge, reflected by childrens' performance in the orthographic lexical decision task.

4.1.2. Does early metacognitive efficiency during orthographic lexical processing predict long term learning improvements in reading-related tasks? If yes, is this mediated by children's early task sensitivity?

A key question beyond the interplay between metacognition and other cognitive systems, is the role of metacognition in regulating one's learning across time. Understanding whether the efficiency of children's metacognitive system predicts the development of their cognitive abilities over time will shed light on the importance of enhancing this ability in classroom and clinical settings.

Only few longitudinal studies have assessed children's language abilities and theory of mind before entering primary school in connection to their metacognition in the memory domain during the early years of primary school (Lecce et al., 2015; Lockl & Schneider, 2007). Both studies revealed a relationship between theory of mind in pre-school ages and metacognition evaluated in the first years of primary school. This relationship was independent of students' language skills. However, to evaluate this relationship these studies were based on second-order false belief and metamemory tasks whose scoring did not control for confidence or type-1 biases of participants.

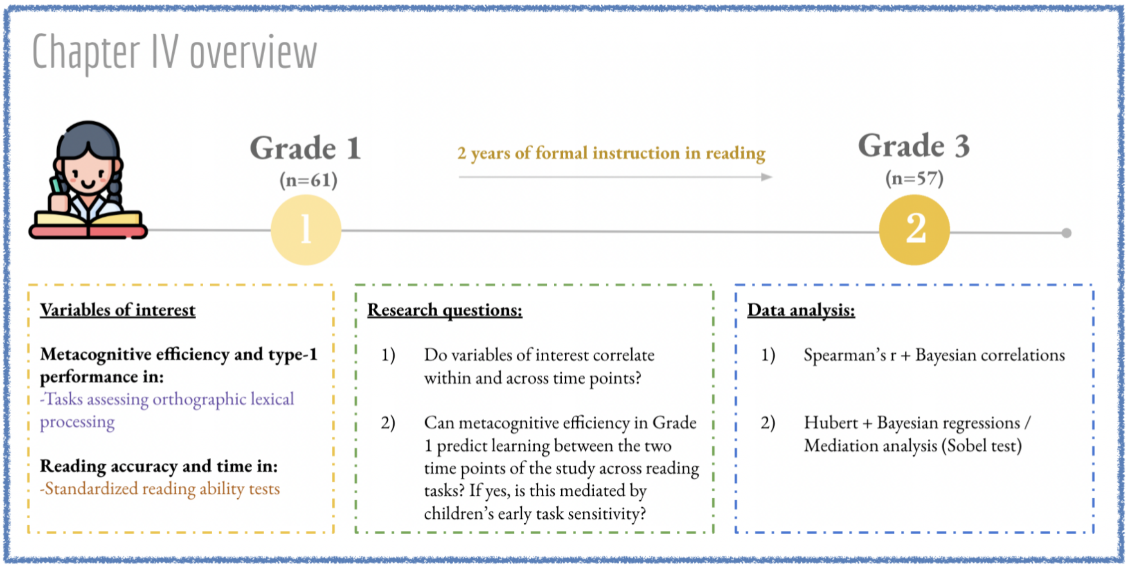
As mentioned in Chapter III, Roebers' et al. (2016) were among the first research groups to assess type-1 and type-2 performance in a spelling task in a longitudinal study including primary school students. In respect to type-1 task performance in the spelling task, they found that it could predict not only participants' future performance in the task, but also future metacognitive performance, suggesting that type-1 performance can be a driving force in the development of metacognition. In Chapters II & III, we provided supporting evidence to this finding, suggesting that type-1 performance in tasks related to orthographic lexical processing in Grade 1 can predict, not only student's longitudinal learning improvements in their type-1 performance, but also longitudinal changes in their type-2 metacognitive ability.

Regarding participants' type-2 metacognitive skill, in the study of Roebers and Spiess (2017) this was not found to predict future spelling performance in the second wave of the longitudinal study (Roebers & Spiess, 2017). However, Rinne and Mazzocco (2014) suggested that type-2 performance in a Problem Verification Test (Murphy & Mazzocco, 2008) measuring mental arithmetic accuracy, is fueling improvements in future type-1 performance (difference in mental arithmetic accuracy between time points) in the same task (Rinne & Mazzocco, 2014). These studies suffer from the lack of control for participants' type-1 performance in the estimation of one's metacognitive skill and confounding effects of type-2 biases.

In the present Chapter, we go beyond the prior work investigating the connection between early metacognitive skills and changes in participant's task performance across time by (i) using bias-free measures of metacognition that also control for type-1 performance and (ii) by providing an objective quantification of type-1 learning skill in tasks directly assessing participants' orthographic knowledge (VA span, lexical decision) and standardized reading ability (reading accuracy, reading time).

We hypothesize that, if metacognition is crucial for the development of children's orthographic lexicon, participants' type-2 metacognitive efficiency in Grade 1 will predict not only improvements in participants' type-1 performance between Grade 1 and Grade 3 within each task, but also students' improvements in the standardized reading ability tests. Moreover, we expect that metacognitive efficiency in the VA span task will significantly predict improvements in participants' type-1 performance in the lexical decision task, reflecting orthographic knowledge.

Finally, if this hypothesis is supported, we aim to explore whether the predictive power of early metacognitive efficiency in participants' learning improvements in the reading-related tasks, is mediated by or if it is independent of students' early task sensitivity.



Variables of interest

Metacognitive efficiency and type-1 performance in:
-Tasks assessing orthographic lexical processing

Reading accuracy and time in:
-Standardized reading ability tests

Research questions:

1) Do variables of interest correlate within and across time points?
2) Can metacognitive efficiency in Grade 1 predict learning between the two time points of the study across reading tasks? If yes, is this mediated by children's early task sensitivity?

Data analysis:

1) Spearman's r + Bayesian correlations
2) Hubert + Bayesian regressions / Mediation analysis (Sobel test)

Figure 18 Schematic representation of Chapter IV research design.

4.2. Methods

4.2.1. Participants

As mentioned in the Chapter of Methodological Considerations, sixty-one children, native Spanish speakers, aged between 6 and 7 years (mean age (±SD): 6.59 ± 0.29, 28 girls), were assessed in the middle of Grade 1 and included in the data analysis of the first wave of the present longitudinal study. The second assessment point of this study occurred 2 years later, in Grade 3, when the sample consisted of fifty-seven children (mean age (±SD): 8.48 ± 0.34, 25 girls). The reduction of the sample was due to outliers identified in the first time point of the study, that were not re-assessed in Grade 3 or participants' change of school. A detailed description of participants' characteristics and process of recruitment can be found in the Chapter Section 1.3.2.1.

4.2.2. Materials and Methods

For the assessment of type-1 and type-2 performance in tasks related to orthographic lexical processing, participants performed a VA span task, assessing the homonymous prerequisite skill of orthographic lexical processing, and a lexical decision task, assessing children's orthographic lexical knowledge. In order to measure participants' standardized reading ability, two standardized reading tests were also administered (word and pseudoword reading) as well as a control task aimed at measuring their non-verbal reasoning abilities (non-verbal IQ). The details of all tasks administered and the assessment process can be found in the Chapter Section 1.3.2.3.

4.2.3. Data analysis

As mentioned in Chapter Section 1.3.2.4., we estimated participants' type-1 performance (d-prime) and type-2 performance (metacognitive efficiency or Mratio) in the reading-related experimental tasks (VA span, lexical decision) under the recently developed hierarchical Bayesian model (Fleming, 2017). In the present chapter, participants' Mratio single-subject estimates were used, which allows for post-fit frequentist analysis including variables from the standardized tests that were not estimated under the hierarchical Bayesian model. In the lexical decision task, participants' mean stimulus presentation duration was used as another measure of type-1 performance, due to the use of the adaptive staircase in this task. As extensively described in the Chapter Section 1.3.2.3., the online adaptive staircase adjusts the presentation time of the stimulus based on participants' discrimination accuracy on each trial, converging to a similar level of performance (i.e., around 70%). Average reading accuracy (%correct of words and pseudowords reading lists) and average reading time (time in sec) were used as standardized reading measures.

First, Spearman's r correlations were used in order to investigate how single-subject estimates of type-2 metacognitive efficiency in the tasks related to orthographic lexical processing associate with participants' performance in the standardized reading tasks (reading accuracy and reading time) and type-1 performance in the rest of the experimental tasks (within and across Grade 1 and Grade 3). For each correlation analysis, False Discovery Rate (FDR) was used for multiple comparison correction and participants' chronological age and intellectual ability were used as covariates.

Second, we investigated whether participants' metacognitive skills in the tasks related to orthographic lexical processing can predict long-term performance improvement within tasks and in the standardized reading tests, using linear and Bayesian regression analyses. For the linear regressions, Huber robust regression was used, which accounts for outliers. Both results are mentioned in the results' section. For the Bayesian regressions, a default prior of 0.354, as implemented in JASP software (van Doorn et al., 2020) was used and the Bayesian inclusion factor ($BF_{inclusion}$) was estimated for every predictor in the model. $BF_{inclusion}$ is calculated by dividing the prior odds of a model including a predictor of interest by the posterior odds (i.e., BF_{10}) excluding this predictor. When $BF_{inclusion} > 1$, it indicates that the model was improved by the addition of this specific predictor.

Finally, in order to investigate whether the predictive value of early metacognition on students' learning improvements is mediated by or is independent of the predictive value of early task sensitivity, we used the online interactive tool "Calculation for the Sobel test" (Preacher & Leonardelli, 2001). Mediation analysis is used to investigate *how and if* the relationship between a predictor independent variable (X: here, early metacognitive efficiency) with an outcome dependent variable (Y: here, learning improvements) is explained by another mediator variable (M: here, early task sensitivity, see: Preacher & Hayes, 2008). The predictive relationship between X and Y, without taking into account the mediator variable, is called the *total effect.* A prerequisite for performing mediation analysis is that the total effect, but also the relationship between X and M, is significant. The relationship between X and Y, when partialling out M, is called the direct effect, while the effect of X on Y through the mediator M, is called the indirect effect. If the direct effect is not significant, this indicates that the effect of X on Y is entirely mediated through the mediator M, and in this case the mediation is *complete.* If both the direct and the indirect effect is significant, that indicates that X has an effect on Y which is mediated through M, but also through other variables and in this case the mediation is *partial.* Finally, if the indirect effect is not significant, this indicates that M does not mediate the effect of X on Y (Preacher & Hayes, 2008). In order to calculate the indirect effect of X on Y, raw regression coefficients from Huber robust regressions were used.

4.3. Results

4.3.1. Descriptives

Descriptive analysis for the measures used to assess participants' type-1 performance in the experimental tasks and task performance in the reading related tasks can be found in Chapter Section 2.3.1. and descriptive statistics regarding participants' type-2 performance in the experimental tasks can be found in Chapter Section 3.3.1.

4.3.2. Correlations between type-2 metacognitive efficiency in the experimental tasks and students' task performance in the reading-related tasks within and across Grade 1 and Grade 3

First, correlations between participants' type-2 metacognitive efficiency in the experimental tasks related to orthographic lexical processing (VA span, lexical decision) and students' task performance on the reading-related tasks (experimental and standardized reading tests) were performed within and across time points of the study. We expected (i) participants with higher type-2 metacognitive efficiency in tasks assessing orthographic lexical processing to exhibit higher type-1 performance on all reading-related tasks within both time points of the study (ii) participants with higher type-2 metacognitive efficiency in the tasks assessing orthographic lexical processing in Grade 1 to exhibit better performance in Grade 3 in the reading-related tasks.

Type-2 metacognitive efficiency and reading performance in Grade 1 and Grade 3

First, no significant correlations were found between type-2 metacognitive efficiency in the experimental tasks related to orthographic lexical processing (VA span and lexical decision) and participants' performance in the standardized reading measures in none of the time points of the study (all ps > 0.45). Bayes factor provides moderate evidence towards the null hypothesis in most of the cases ($BF_{10} < 0.33$, see Table 11), except in the case of the negative association between participants' metacognitive efficiency in the lexical decision task and reading accuracy in Grade 1 ($BF_{10} = 6.579$), in which moderate evidence towards the alternative hypothesis was provided.

Second, in Grade 1 we found some significant or marginally significant negative associations between participants' type-2 metacognitive efficiency and type-1 d-prime within the tasks assessing orthographic lexical processing (VA span: $r = -0.329$, $p = 0.048$, $BF_{10} = 19.171$; lexical decision: $r = -0.307$, $p = 0.070$, $BF_{10} = 35.759$, FDR-corrected p-values), but also between students' metacognitive efficiency in the lexical decision task and type-1 performance in the VA span task ($r = -0.303$, $p = 0.094$, $BF_{10} = 1.721$, FDR-corrected, see Table 11). As noted in the parenthesis, Bayesian analysis provided anecdotal to strong evidence towards the alternative hypothesis in all of these associations. On the contrary, in Grade 3 no association was found between participants' type-2 metacognitive efficiency and type-1 performance in any of the experimental tasks (all $ps > 0.5$, see Table 11).

4.3.3. Investigating the relationship between participants' early metacognitive efficiency in the tasks related to orthographic lexical processing and the developmental changes in participants' task performance in the reading-related tasks between Grade 1 and Grade 3

We first examined whether metacognitive efficiency in Grade 1 in the VA span task (prerequisite of orthographic lexical processing), can predict students' learning improvements between the two time points of the study within this task, improvements in the lexical decision task, reflecting orthographic knowledge, and in the standardized tests (reading accuracy and time), reflecting visual word recognition abilities of the students. Second, we tested whether metacognitive efficiency in the lexical decision task, can predict students' learning improvements within the same task, and in the standardized reading tests. Next, in the cases that metacognitive efficiency was a significant predictor of learning across time, we investigated whether this predictive value is independent of, or mediated by, participants' type-1 performance in Grade 1.

Regression analyses were performed to predict these longitudinal performance changes, measured as the difference in task performance between Grade 1 and Grade 3 within each task, using metacognitive efficiency in the experimental reading-related task, age and non-verbal IQ at Grade 1 as predictors. Analyses performed using a robust to outliers regression model (Hubert regression) and Bayesian regression models showed that:

a) metacognitive efficiency in the VA span task in Grade 1 was a marginally significant positive predictor of longitudinal changes in the type-1 d-prime within the same task ($p = 0.061$, $BF_{10} = 1.956$, see Table 12 and Figure 19a), but did not predict significantly participants' learning improvements in the lexical decision task and the standardized reading test (all $ps > 0.615$).

Table 11 Spearman's correlations between metacognitive efficiency (Mratio) in the reading-related experimental tasks and a) students' performance on standardized tasks measuring reading ability, b) type-1 task sensitivity in the rest of the experimental tasks, in Grade 1 and Grade 3 (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, FDR corrected). Correlations were controlled for participant's age and intellectual ability (non-verbal IQ, Matrices-WISC). Significant correlations are noted in bold font.

	Standardised reading tasks				*Linguistic tasks related to orthographic lexical processing*					
	Reading Accuracy (%correct)		Reading Time (sec)		VA Span d-prime		Lexical Decision d-prime		Lexical Decision Mean stimulus duration (msec)	
	Grade 1	**Grade 3**	**Grade 1**	**Grade 3**	**Grade 1**	**Grade 3**	**Grade 1**	**Grade 3**	**Grade 1**	**Grade 3**
VA Span Mratio Grade 1	r = 0.009 p = 0.997 BF_{10} = 0.222	r = 0.171 p = 0.482 BF_{10} = 0.278	r = 0.050 p = 0.856 BF_{10} = 0.180	r = -0.179 p = 0.451 BF_{10} = 0.199	**r = -0.329** **p = 0.048** **BF_{10} = 19.171**	r = -0.004 p = 0.997 BF_{10} = 0.184	r = 0.015 p = 0.997 BF_{10} = 0.176	r = -0.116 p = 0.659 BF_{10} = 0.200	r = 0.096 p = 0.734 BF_{10} = 0.187	r = -0.086 p = 0.764 BF_{10} = 0.178
VA Span Mratio Grade 3	r = 0.013 p = 0.997 BF_{10} = 0.175	r = 0.071 p = 0.789 BF_{10} = 0.279	r = 0.118 p = 0.658 BF_{10} = 0.301	r = 0.004 p = 0.997 BF_{10} = 0.207	r = -0.192 p = 0.416 BF_{10} = 0.415	r = 0.071 p = 0.789 BF_{10} = 0.184	r = -0.135 p = 0.635 BF_{10} = 0.223	r = 0.151 p = 0.566 BF_{10} = 0.461	r = -0.022 p = 0.997 BF_{10} = 0.198	r = -0.094 p = 0.745 BF_{10} = 0.256
Lexical Decision Mratio Grade 1	r = -0.136 p = 0.580 BF_{10} = 6.579	r = 0.081 p = 0.764 BF_{10} = 0.366	r = 0.080 p = 0.764 BF_{10} = 0.217	r = -0.169 p = 0.482 BF_{10} = 0.211	**r = -0.303** **p = 0.094** **BF_{10} = 1.721**	r = -0.133 p = 0.635 BF_{10} = 0.294	**r = -0.307** **p = 0.070** **BF_{10} = 35.759**	r = -0.007 p = 0.997 BF_{10} = 0.200	r = 0.129 p = 0.612 BF_{10} = 0.921	r = -0.104 p = 0.709 BF_{10} = 0.211
Lexical Decision Mratio Grade 3	r = 0.084 p = 0.764 BF_{10} = 0.173	r = 0.071 p = 0.789 BF_{10} = 0.443	r = -0.120 p = 0.648 BF_{10} = 0.390	r = -0.016 p = 0.997 BF_{10} = 0.982	r = -0.056 p = 0.846 BF_{10} = 0.192	r = 0.083 p = 0.764 BF_{10} = 0.262	r = 0.010 p = 0.997 BF_{10} = 0.180	r = -0.010 p = 0.997 BF_{10} =0.181	r = 0.014 p = 0.997 BF_{10} = 0.182	r = -0.128 p = 0.633 BF_{10} = 1.815

b) metacognitive efficiency in the lexical decision task in Grade 1 was indicated as a significant positive predictor of longitudinal changes in the type-1 d-prime within the same task ($p = 0.032$, $BF_{10} = 1.092$, see Table 13) and of participants' longitudinal changes in their standardized reading accuracy ($p < 0.001$, $BF_{10} = 2.620$, see Table 13 and Figure 19b), but not in the standardized reading time ($p = 0.180$, $BF_{10} = 0.426$).

Finally, in order to examine whether the effect of early metacognitive efficiency on learning improvements in reading related variables was mediated by participants' early task sensitivity, we fit mediation analysis models for the regression models in which metacognitive efficiency significantly predicted learning (prerequisite in order to perform mediation analysis, see Figure 20). Early task sensitivity in the lexical decision task was found to mediate the effect of early metacognitive efficiency in the same task on students' learning improvements in the lexical decision task and their standardized reading accuracy (see Figure 20).

Table 12 Regression analysis (Hubert regression) of participants' longitudinal changes in task performance between the two time points of the longitudinal study in a) the VA Span task (n=47), b) the lexical decision task (n=48), c) the standardized reading tests (n=53), with metacognitive efficiency in the VA span task, age and non-verbal IQ at Grade 1 as predictors. BFinclusion factor represents the change from prior to posterior probabilities of a model when a predictor is added in the equation (BFinclusion > 1 indicates that the predictor improves the model).

	d-prime changes in VA Span task (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
VA Span Mratio (Grade 1)	0.110	1.903	**0.060**	**1.956**	**2.537**
Age	-0.183	-1.023	0.315	0.604	1.000
Non-verbal IQ	-0.027	-1.611	0.110	1.132	1.000

	d-prime changes in Lexical Decision task (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
VA Span Mratio (Grade 1)	-0.034	-0.430	0.679	0.420	0.591
Age	0.113	0.456	0.659	0.387	1.000
Non-verbal IQ	-0.042	-1.836	0.071	0.887	1.000

	Reading Accuracy (%correct) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	Bfinclusion (Age, IQ in null model)
VA Span Mratio (Grade 1)	-0.015	-0.509	0.615	0.361	0.676
Age	-0.053	-0.576	0.564	0.299	1.000
Non-verbal IQ	-0.012	-1.454	0.150	0.530	1.000

	Reading Time (msec) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
VA Span Mratio (Grade 1)	0.844	0.061	0.951	0.231	0.460
Age	22.478	0.521	0.609	0.244	1.000
Non-verbal IQ	6.099	1.564	0.126	0.352	1.000

Table 13 Regression analysis (Hubert regression) of participants' longitudinal changes in task performance between the two time points of the longitudinal study in a) the lexical decision task (n=48) and b) the standardized reading tests (n=53), with metacognitive efficiency in the lexical decision task, age and non-verbal IQ at Grade 1 as predictors. $BF_{inclusion}$ factor represents the change from prior to posterior probabilities of a model when a predictor is added in the equation ($BF_{inclusion} > 1$ indicates that the predictor improves the model).

	d-prime changes in Lexical Decision task (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Lexical Decision Mratio (Grade 1)	0.119	2.119	**0.032**	**1.092**	**1.557**
Age	0.298	1.461	0.149	0.676	1.000
Non-verbal IQ	-0.014	-0.764	0.462	0.647	1.000

	Reading Accuracy (%correct) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Lexical Decision Mratio (Grade 1)	0.083	3.514	**<0.001**	**2.620**	**5.250**
Age	0.041	0.482	0.633	0.413	1.000
Non-verbal IQ	0.001	0.159	0.874	0.413	1.000

	Reading Time (msec) changes in PROLEC standardized tests (Grade 3-Grade 1)				
	β	t	p	BFinclusion	BFinclusion (Age, IQ in null model)
Lexical Decision Mratio (Grade 1)	-16.751	-1.353	0.180	0.426	0.908
Age	18.362	0.403	0.695	0.272	1.000
Non-verbal IQ	3.193	0.759	0.452	0.256	1.000

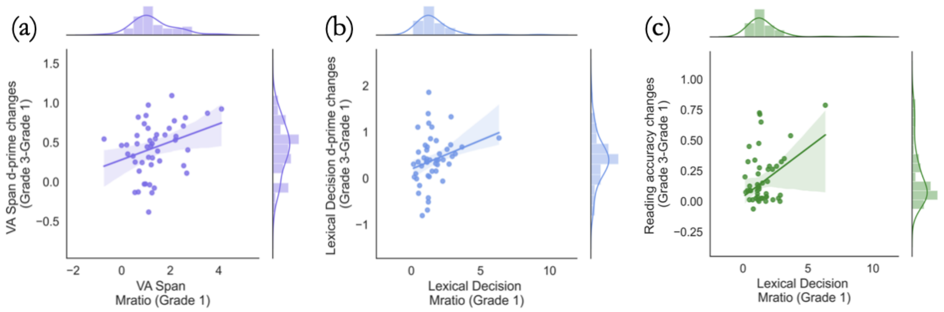


Figure 19 Linear relationship between longitudinal changes in participants' task sensitivity in reading-related tasks and Mratio in Grade 1 (a: VA span task (task performance long. changes-Mratio in Grade 1), b: lexical decision task (task performance long.changes-Mratio in Grade 1), c: long. changes in reading accuracy - Mratio in lexical decision task).

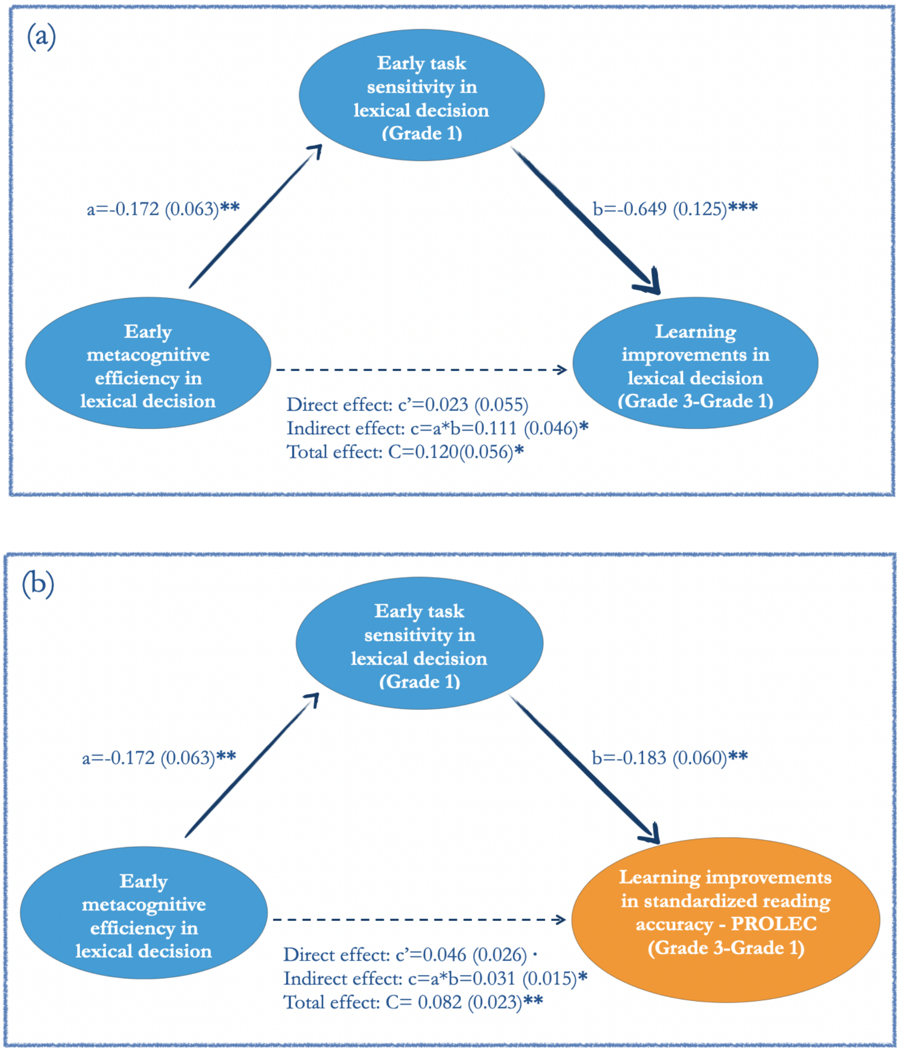


Figure 20 Mediation analysis showing the mediating role of early task sensitivity in the lexical decision task on the effect of early metacognitive efficiency in the same task on students' learning improvements: a) in the lexical decision task (n=48), b) in the standardized reading accuracy test (n=47) (p-value: $p < 0.01$., $p < 0.05$*, $p < 0.01$**, $p < 0.001$***).

4.3.4. Discussion

The main goals of Chapter IV were two-fold. First, we wanted to explore whether students' metacognitive ability is related to reading in different stages of primary school, while focusing on the study of skills tapping into orthographic lexical knowledge. Second, we wanted to investigate whether early metacognitive efficiency can predict long term changes in students' performance in tasks related to orthographic lexical processing during the first years of primary school. Below, we briefly report the main results that will be discussed.

First, there was no significant evidence for any association between metacognitive efficiency on the experimental tasks related to orthographic lexical processing and reading performance in the standardized reading tests, neither in earlier (Grade 1) nor in later stages of reading acquisition (Grade 3). Interestingly, in Grade 1, we found some negative correlations, supported by evidence provided in the bayesian analysis, between participants' metacognitive efficiency and type-1 task performance (d-prime), both within the tasks assessing orthographic lexical processing (VA span, lexical decision) and between metacognitive efficiency in the lexical decision task and type-1 performance in the VA span task. Moreover, participants' metacognitive efficiency in the lexical decision task negatively correlated with standardized reading accuracy in the same time point of the study. Last, we showed that early metacognitive efficiency in the VA span task was indicated as a marginally significant predictor of students' longitudinal improvements in task performance within the same task, while early metacognitive efficiency in the lexical decision task, could significantly predict longitudinal improvements in students' performance both within this task and in participants' standardized reading accuracy. Mediation analysis models indicated that this effect of early metacognitive efficiency on student's longitudinal learning in reading related tasks was mediated by early task sensitivity.

Metacognitive ability and reading performance in earlier and later stages of reading acquisition

Previous research investigating the relationship between metacognitive ability and performance in standardized tests across different domains reported mixed findings. Positive associations have been reported in several domains, like mathematics (Bellon et al., 2019), emotion recognition (Kelly and Metcalfe 2011), spelling, and text comprehension (Griffin et al. 2008). However, other studies reported no association between metacognitive monitoring and cognitive ability such as memory strategies (Kelly et al., 1976) and text comprehension skills (Griffin et al., 2009). It has been suggested that these results may be attributed to the use of metacognitive indexes which are susceptible to the confounding effects of confidence bias, participants' level of type-1 performance, and methods that permit guessing, which can differentially affect the estimation of metacognitive accuracy in high vs poor performers (Vuorre & Metcalfe, 2021). In the present study, we used a bias-free signal detection theoretic framework, including Bayesian estimation, to assess participants' metacognitive efficiency, overcoming these issues.

In Chapter II, we observed significant correlations in type-1 performance in the experimental tasks related to orthographic lexical processing (VA span and orthographic lexical decision), both across these tasks and with participants' performance in the standardized tests measuring reading accuracy and reading time, in line with previous research linking tasks assessing orthographic lexical processing with standardized reading performance (Ginestet et al., 2021; Valdois et al., 2019). However, contrary to our predictions, no significant association was found in frequentist correlational analysis between participants' metacognitive efficiency in any of the experimental tasks and their performance in the standardized reading tests in any of the time points of the study.

Taking into account the neurocognitive changes happening in this critical age window, one possibility is that the neurodevelopmental trajectories of the systems that are relevant for acquiring reading are somehow segregated from the systems that support attention and cognitive control, and hence, metacognition in the early stages of reading acquisition. Recently, Filevich et al. (2020) have suggested that the age of 6 (like our children participants in the first branch of this longitudinal study) is a critical age in the development of metacognitive monitoring. Using tasks in which children had to recognize and report their knowledge certainty in the task, they showed that children's ability to correctly identify and explicitly report that they did not know is associated with key changes in cortical thickness in the medial orbitofrontal cortex (Filevich et al., 2020). Additionally, Brod et al. (2017) suggested that the first year of schooling brings a shift in children's cognitive abilities that may be critical for metacognition. Specifically, during this first year of schooling, students between 5 and 6 years of age showed great improvements in tasks requiring executive control functions, which are also linked to activity changes in parietal cortex regions associated with attention control (Brod et al., 2017).

Based on the above considerations, one would expect that following the initial, early development of metacognition and reading skills, the pattern of results would change in the second branch of this longitudinal study when children have already received 2 years of formal instruction in reading, which was not the case in our study. However, in Chapter III, we found that metacognitive efficiency in all reading-related tasks significantly decreased in later stages of reading acquisition. An explanation given was that childrens' experience on the task and their improvement on type-1 performance could have reduced the need of employing metacognitive resources to control their performance. This may have also affected the relationship between metacognitive efficiency and standardized reading ability measures in Grade 3.

Another factor that needs to be taken into account when interpreting the current results is the outburst of the pandemic. Students participating in this longitudinal study, received one full year of reading instruction in Grade 1, but during Grade 2 their learning process was interrupted by the pandemic. Hence, one needs to consider that when we re-tested the same kids in Grade 3 each kid was differentially affected in their reading development and development of metacognitive skills by their homeschooling conditions i.e., time spent by their parents to support students' learning, availability of devices giving access to online classes etc. A further cross-sectional investigation in a control group with no previous experience with the tasks or investigation of this research question in later stages of primary school may be useful to determine the factors that mediate the interplay between metacognition and general reading ability.

Finally, intriguingly, Bayesian analysis provided evidence that metacognitive efficiency in the lexical decision task negatively correlated with type-1 performance both within the task, but also with type-1 performance in the VA span and participants' standardized reading accuracy in Grade 1. Moreover, participants' metacognitive efficiency in the VA span task negatively correlated with their type-1 performance within this task. This pattern of results was not observed in the second time point of our study (Grade 3). We suggest that at this early age of development, students with lower type-1 performance may compensate for their difficulties by means of an increase in metacognitive monitoring ability, possibly driven by increased signaling from error monitoring systems which would lead to them knowing better when they are incorrect or feel uncertain about their decisions. Accordingly, it has been suggested that the development of metacognitive monitoring in the early ages of primary school is particularly related to the efficient monitoring of incorrect responses (Destan et al., 2014). To support this hypothesis, we ran some additional post hoc correlation analyses between participants' type-1 task sensitivity and the percentage of incorrect responses rated with low confidence on the task. In Grade 1, we found that type-1 sensitivity both in the lexical decision task and the VA span task negatively correlated to the proportion of incorrect responses rated with low confidence within each task, so that children with higher rates of this proportion had lower type-1 performance (r = -0.279, p = 0.066 and r = -0.419, p = 0.005 respectively); likewise, type-1 sensitivity in the VA span task negatively correlated to the percentage of low confidence incorrect responses in the lexical decision task (r = -0.345, p = 0.035). Moreover, we found that the percentage of incorrect responses rated with low confidence in the lexical decision task was negatively, but not significantly, correlated with participants' performance in both of the standardized tests measuring reading ability (reading accuracy: r = -0.229, p = 0.123, reading time: r = 0.243, p = 0.109). In Grade 3, no significant correlations were found between those variables.

One possible explanation regarding the negative correlation between participants' type-1 performance in the VA span task and the standardized reading measures with error monitoring indexes in the orthographic lexical decision task in Grade 1 is that early readers that already have in place the necessary tools for fluent reading -such as VA span skills (Valdois et al., 2019) may use more implicit or automatic ways of reading strategies (automatic sight word reading), having less need of monitoring their performance at this stage of reading acquisition. Conversely, students who exhibit lower reading performance and orthographic knowledge do this in a more controlled fashion and become more able to detect their errors efficiently in reading-related tasks.

Metacognition and long-term learning during early childhood

Interestingly, despite the observed lack of associations between metacognitive ability and reading performance within earlier and later stages of reading acquisition, we found that students' metacognitive efficiency in Grade 1 in some linguistic tasks could predict participants' performance improvement between Grade 1 and Grade 3. Early metacognition in VA span task could marginally predict long-term improvements of participants' type-1 task sensitivity within this task, while early metacognition in lexical decision task could significantly predict long-term improvements of participants' type-1 task sensitivity within the task, but also improvements in participants' standardized reading accuracy.

Metacognition has been long considered as a driving force at regulating individuals' learning, by monitoring uncertainty, guiding exploration and controlling performance (Flavell, 1979; Metcalfe, 2009; Narens, 1990). In educational practice, it has been suggested that metacognition can regulate study time allocation for easy vs hard tasks, or direct students' need for information seeking or assistance (Desender et al., 2018; Dunlosky et al., 2021; Son & Metcalfe, 2000). Of note, most of these studies investigating the link between metacognition and learning have focused on associations within a certain time point. Few studies have investigated the relationship of metacognitive monitoring and type-1 performance longitudinally like the present study. Roebers and Spiess (2016) longitudinally tracked the development of online metacognitive monitoring in early primary school in the spelling domain but did not observe that metacognitive monitoring at the beginning of the study (children's age: 7 y.o.) predicted children's performance in a spelling task 8 months later in Grade 2. These results are in contrast to Rinne and Mazzocco's (2014) study showing that early metacognitive skills can predict long-term improvements in performance in an arithmetic task three years later in primary school (Grade 5 to Grade 8). Differences among studies, including ours, may be attributed to the fact that here and in Rinne and Mazzocco's study, the link between metacognition and learning was assessed based on the change in performance across two time points. For example, Roebers & Spiess (2017) merely correlated type-2 metacognitive sensitivity at time point 1 with type-1 performance at time point 2 (Roebers & Spiess, 2017) without quantifying any change in performance across time points as we have done here. Our observation that metacognitive skill is predictive of subsequent learning effects across time is in line with prior educational studies suggesting that individuals' monitoring ability of their own performance is fundamental for learning (Metcalfe & Kornell, 2007; Rawson et al., 2011).

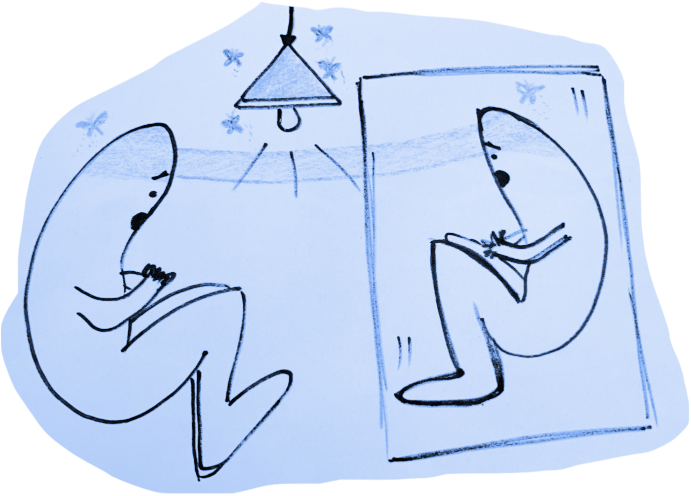
In Chapter II, in the same tasks that we here report that early metacognitive efficiency was indicated as a significant predictor of learning, we showed that early task performance could negatively predict learning improvements across time, indicating that students with lower reading ability in Grade 1 improve more during the first years of primary school, catching up with children who entered the primary school being more fluent in reading. Moreover, in Grade 1 we noted some negative correlations between students' metacognition in a reading task and their task performance in the same or another reading task. To follow up these findings in the cases in which early metacognitive performance was indicated a significant predictor of task performance learning improvements across time, we performed a mediation analysis, which indicated early task sensitivity as a mediator on this effect.

Taken together, these results indicate that students who are less experienced in reading in the early stages of reading acquisition rely more on metacognitive strategies and error monitoring processes in order to catch up with their peers, and stop doing so in later stages of reading acquisition when the gap between inexperienced and fluent readers is decreasing. Hence, the co-evaluation of early task sensitivity and metacognitive efficiency in the first year of primary school may act as a marker of whether students with lower reading at this stage will manage to catch up with their peers, or if they are at risk of developing a reading disorder.

Chapter IV Summary

- Overall, using frequentist statistics, no associations were found between metacognitive efficiency in tasks assessing orthographic lexical processing (VA span, lexical decision) and students' standardized reading performance (reading accuracy and time) neither in Grade 1, nor in Grade 3. However, Bayesian analysis provided evidence towards a negative relationship between participants' metacognitive efficiency in the lexical decision task in Grade 1 and students' task performance in the VA span task and in standardized reading accuracy at the same time point of the study.
- Students' with higher early metacognitive efficiency in the lexical decision task in Grade 1 were the ones who showed the highest learning improvements in their type-1 performance across time, both within this task and in standardized reading accuracy. In Chapter II, we showed that the highest learning improvements in the reading-related tasks were shown by the students who exhibited the lowest early task sensitivity. Mediation analysis indicated that the effect of early metacognitive efficiency on learning improvements in these tasks, is mediated by early task sensitivity.
- Taking together these two findings, we suggest that students who begin Grade 1 with higher reading abilities, and have mastered their orthographic mappings, may have automatized their visual word reading skills (Ehri, 2005; Verhoeven & Perfetti, 2017), and hence, have less need of monitoring their performance during sight-word reading. On the contrary, students that are less fluent in reading in the first year of primary school, may rely more on metacognitive resources, in order to carefully monitor their performance and build their orthographic lexicon through analytic reading.

5. Chapter V: General Discussion



5.1. Summary of findings and general conclusions

The current dissertation used a within-subject, longitudinal design to assess the students' metacognitive monitoring ability during the first years of reading acquisition. Specifically, this thesis aimed to investigate whether metacognition influences the development of individuals' orthographic lexical processing and visual word recognition, which are crucial skills in the transition from stages of reading through decoding, towards consolidated fluent reading stages (Ehri, 2014; Frith, 1985).

To this end, we systematically assessed students' metacognitive ability and task performance in Grade 1 (6-7 years old) and, anew, in Grade 3 (8-9 years old) in the following tasks: a) a VA span task, estimating individuals' ability to process in parallel multiple strings with one eye fixation, which has been considered a prerequisite skill for orthographic lexical processing (Bosse et al., 2007; Bosse & Valdois, 2009; Lallier et al., 2013), b) a lexical decision task, which assesses the efficiency of retrieving whole orthographic forms from the orthographic lexicon (Stanovich, 1993) and c) an emotion recognition task, which is a non-linguistic task and was used for comparative purposes, in order to clarify whether effects observed were reading-related or not. The index of metacognitive efficiency (Mratio) was used to estimate participants' metacognitive ability, which was calculated under a Bayesian Signal Detection Theory model (Fleming, 2017). Moreover, in both assessment points students' average standardized reading accuracy and time was measured in lists of words and pseudowords, reflecting individuals' visual word recognition skills (Cuetos, Rodrigues, Ruano, 1996).

Understanding how metacognitive monitoring develops over time in connection with the abovementioned skills during the first years of reading acquisition is pertinent to our better understanding of the higher order thinking skills involved in guiding learning and specifically in the development of the reading system and orthographic processing skills, before fluent reading evolves. Gaining insight into these processes during the first years of primary school can have important implications in defining when it is beneficial for students' to reflect on their own knowledge and performance, and hence, in designing educational programs for training metacognition to assist efficient sight word reading and learning in distinct domains.

Below, we summarize the main findings related to each of the main goals set for the present study, place them in the context of existing literature and suggest future research directions:

5.1.1. The development of metacognition in tasks related to orthographic lexical processing (vs a non-linguistic task) during the first years of primary school.

To address our first research question, regarding how metacognition develops during the first years of primary school, we compared students' metacognitive efficiency in tasks assessing orthographic lexical processing and non-linguistic tasks mentioned above, between the two time-points of the study (Grade 1 and Grade 3). Also, we evaluated and compared participants' type-1 task performance (d-prime, stimulus presentation duration) and other type-2 performance indexes (mean confidence, metacognitive sensitivity).

Interestingly, we found that students' metacognitive efficiency decreased significantly over time across linguistic and non-linguistic tasks, despite increases in type-1 performance. Moreover, we showed that there was a decrease in the other measures of type-2 performance between time points of the study (i.e., reduction of the overconfidence bias). As extensively mentioned in the General Introduction and Chapter III, metacognitive monitoring has been considered to develop in its explicit form mainly after 5 years of age and increase during childhood (Schneider, 2015; Destan et al., 2014; Rohwer et al., 2012). However, most of these studies use metacognitive indexes which do not avoid the confounds of confidence bias and type-1 bias. Metacognitive efficiency, a recently developed index that accounts for these biases, has been scarcely studied during development. Hereby, there are very few studies studying metacognitive efficiency in the human lifespan. A very recent study from Moses-Payne et al. (2021) has suggested that metacognitive efficiency increases in the transition from childhood (8-9 years old) to adolescence (12-13 years old) and remains stable through late adolescence (16-17 years old) in a visual perceptual task (Moses-Payne et al., 2021), while Palmer et al. (2014) had previously found that metacognitive efficiency decreases by approx. 0.6% every year of adult life (18-64 years old, see: Palmer et al., 2014). Notwithstanding, these studies used a cross-sectional design and they both raised the necessity for studying metacognitive efficiency longitudinally during development to avoid inter-individual variability across the different developmental stages. Bang et. al. (2019), to our knowledge, were the first ones to assess metacognitive efficiency using a within-subject design in adults performing a visual perceptual task over 7 sessions in distinct days, and showed that metacognitive efficiency decreases with repeated experience, while individuals become more competent in type-1 performance (Bang et al., 2019). In this study, we tracked metacognitive efficiency longitudinally during the first years of primary school.

When interpreting the results of a longitudinal design, one needs to take into account that individuals gain experience on the tasks, but also develop the related skills assessed across time. In the current study, in the lexical decision and the emotion recognition task we adapted the stimulus presentation duration for each individual to avoid ceiling effects of accuracy in both time points of the study, while in the VA span task, we followed previous literature suggesting that students in these stages of primary school attain accuracy scores of 60-70% (Lallier et al., 2016). However, we showed that metacognitive efficiency decreased, while task performance increased across time, despite the above-mentioned manipulations.

We propose that the observed decrease in students' type-2 metacognitive efficiency may be attributed to students' increased experience and the adoption of a more automated processing in the type-1 task which does not demand as much the employment of metacognitive resources, and not to a change in metacognitive ability *per se* due to development. The idea that type-2 metacognitive and type-1 perceptual decisions depend upon common neural resources is not new. Maniscalco and Lau (2017) suggested that regions in the anterior prefrontal cortex (aPFC) contribute both to visual metacognition and to perceptual vigilance (the ability to stay attentive on the perceptual type-1 task), and that reducing metacognitive demands can increase perceptual vigilance (Maniscalco et al., 2017). Shekhar and Rahnev (2018) later suggested that aPFC plays a crucial role in an individual's metacognitive ability by combining perceptual information (strength of sensory evidence) with non-perceptual information (Shekhar & Rahnev, 2018). Interestingly, both of these studies showed that there was a decrease in metacognitive efficiency in the second half of task performance, which had been mainly attributed to fatigue (Maniscalco et al., 2017). However, in the literature, one can find a variety of factors that have been proposed to be detrimental to metacognition and that arise with time spent on a type-2 task, such as confidence leak - i.e., confidence on a given trial or task influencing individuals' confidence on following trials or tasks- (Rahnev et al., 2015), action biases - i.e., choosing one response over another because of preferred action movement - (Fleming et al., 2015) and arousal effects (Allen et al., 2016). Therefore, we suggest that these factors may also lead to the increase of participants' metacognitive noise longitudinally (see Discussion in Chapter III, Bang et al., 2019), which together with the familiarization of the participant with the task, favours a more automated mode of performance in the tasks. We suggest that this process, enhanced by the natural development of skills across time (e.g., development of sight word reading instead of decoding / increase of one's visual attention span during development) which also contributes to the decrease of type-1 sensory noise, may be related to the observed significant decreases in metacognitive efficiency during the first years of primary school.

Finally, we would also like to draw researchers' attention in a very recent work that was published after the completion of the present work, showing a robust small but positive correlation between individuals' metacognitive efficiency and confidence bias (mean confidence) in certain tasks using large-scale datasets from the Confidence Database (Xue et al. 2021; Rahnev et al. 2020). We hence also consider here the possibility that the childrens' decrease in confidence bias over time in our study (see Figure 14b), may have influenced the decrease in metacognitive efficiency that we observed across tasks. To clarify this, we performed Spearman's r correlations between changes in the metacognitive efficiency and changes in confidence bias between Grade 1 and Grade 3 within each task. Only in the emotion recognition task a positive relationship was observed between these variables (VA Span: $r = -0.118$, $p = 0.664$, lexical decision: $r = 0.016$, $p = 0.918$, emotion recognition: $r = 0.308$, $p = 0.106$), suggesting that the observed decrease in metacognitive efficiency in the VA and lexical decision tasks is unlikely to be driven by confidence bias changes. We note also that the present thesis used the optimal existing methods to date to measure metacognition, while mitigating the confounds of type-1 and confidence biases. Notwithstanding, we consider it crucial for future studies to take into account Xue et al. 's recent study (2021) which raises the possibility of further improving the current methods to measure metacognitive ability.

5.1.2. The use of domain-general or domain-specific mechanisms supporting metacognition in tasks related to orthographic lexical processing and a non-reading task during early childhood.

The second goal of the present study was to shed light on the nature of the mechanisms used to support metacognition in the first years of primary school. To address this question, we examined whether metacognitive efficiency associates across tasks within Grade 1 and Grade 3, and if the magnitude of the correlation changes across time. In Grade 1, we found a strong positive correlation between students' metacognitive efficiency in the lexical decision and the emotion recognition task, using group-level estimates of metacognition under the hierarchical Bayesian model (Fleming, 2017), which account for single-subject estimates with high uncertainty on the group, and has been considered more accurate for estimating covariance across tasks. When associating single-subject estimates using frequentist statistics at the same time point of the study, a weaker significant correlation was also found between the lexical decision and the VA span task. Notwithstanding, no association was found between any pair of tasks in Grade 3.

The few existing studies examining the issue of domain generality/specificity during development have suggested that there is a shift from domain specific to domain general mechanisms in middle childhood (Bellon et al., 2019; Geurten et al., 2018; Vo et al., 2014). However, these studies diverge on the suggested developmental stage at which the shift is occurring, and on the tasks/domains studied. For instance, Geurten et al. (2018) and Vo et al. (2014) highlight that until the age of 8 there are no detected domain general mechanisms supporting metacognition, while Bellon et al. (2019) place the timeline of shift towards domain generality at this age. In our study, we found little evidence towards domain general mechanisms already from Grade 1 (6-7 years old). However, contrary to our expectations, not only did we not find such evidence in Grade 3, but also the magnitude of correlation found in Grade 1 between the lexical decision and the emotion recognition task decreased significantly.

Interpreting this pattern of results is not straightforward. However, our study differs substantially methodologically from previous developmental studies in the following points: a) existing studies use a cross-sectional design, while here a longitudinal within-subject design was used and b) unlike the above-mentioned studies, we have used a metacognitive index (Mratio) which accounts for confidence and type-1 biases.

We propose that the pattern of results may be linked: a) to the different strategic cues students use to provide explicit confidence ratings across development and b) to the extended experience participants had on these tasks and the natural development of the skills assessed during primary school (i.e., orthographic lexical processing skills turn from immature to mature between the two time points of assessment). When a skill assessed is developing, or when students are assessed in a certain task they have no experience in, they may use cues that depend more on peripheral characteristics of the task (e.g., how fast the stimulus appears) than on their knowledge on the developing skill, or they may rely more on low-level generic feelings of confidence.

Morales et al. (2018), in an fMRI study in adults, suggested that domain-general and domain-specific mechanisms co-exist in the human prefrontal cortex in a perceptual and a memory task, even when domain-general mechanisms are not reflected on behavioural measures. They suggested that the aPFC holds content-rich, domain-specific metacognitive representations, while more generic feelings of confidence are represented in a widespread frontal and posterior network (Morales et al., 2018). As we mentioned in the previous section of the General Discussion, activity in the aPFC has been linked with individuals' metacognitive efficiency measures. Also, the aPFC may play a role in combining non-perceptual factors with generic feelings of confidence (Shekhar & Rahnev, 2018).

We suggest that as individuals gain experience in a certain task, different non-perceptual factors begin to play cumulative effects within a particular domain, leading to more specific domain mechanisms supporting metacognitive efficiency across the different tasks (here in Grade 3 students). It would be interesting for future neuroimaging studies, to assess how and if domain-general and domain-specific patterns of brain activity supporting metacognition change when pairs of tasks are assessed in a longitudinal design as participants gain more experience with the tasks. Moreover, it would be useful to examine whether behavioural measures of metacognitive efficiency alter when providing explicit generic training to participants on which cues to use to provide confidence judgments in a perceptual task.

Understanding whether metacognition can be boosted and trained holistically or whether it lays on specific mechanisms to the domain studied, and whether this pattern changes with age, can have important implications in designing educational programs training metacognitive efficiency.

5.1.3. The role of metacognition in regulating students' learning, namely, during the development of students' orthographic lexicon and reading ability.

The third and most important goal of this study was to elucidate the role of metacognitive monitoring in early reading acquisition and especially regarding the development of orthographic lexical processing. Towards this end, we examined whether metacognitive efficiency in the VA span and the lexical decision task associate with students' type-1 performance in these tasks, and in tasks measuring individuals' standardized reading ability in Grade 1 and Grade 3. Moreover, we tested whether early metacognitive efficiency and early task performance can independently or in conjunction predict changes in reading performance across time.

We found that in Grade 1 metacognitive efficiency negatively correlated with task performance within the tasks assessing orthographic lexical processing, but also that metacognitive efficiency in the lexical decision task negatively correlated with task performance in the VA span and participants' standardized reading accuracy in the same time point (the latter correlation was only supported by Bayesian statistics). No association was found between metacognitive efficiency and reading performance in Grade 3. Interestingly, we separately showed that early metacognitive efficiency and early type-1 task performance in the lexical decision task can predict improvements in task performance across time, both in the lexical decision task, which reflects orthographic lexical access, and in participants' standardized reading accuracy measure, reflecting sight word reading. A mediation analysis revealed that the effect of early metacognitive efficiency on students' learning improvements was mediated fully or partially by their early task sensitivity.

Taken together, these findings suggest that, in Grade 1, students who are struggling more when performing the reading related perceptual type-1 tasks (VA span, lexical decision, standardized reading accuracy) are the ones who employ more metacognitive resources to monitor their performance and detect their errors in the task assessing orthographic lexical access (lexical decision task). According to the MTM model (Ans et al., 1998), presented in the Introduction, we propose that students' with lower visual attention skills, or in a less advanced reading stage (e.g., partial alphabetic stage (Ehri, 1995; Frith, 1985), use analytic reading to access or strengthen word traces in their episodic memory, which requires more attentional resources and error monitoring skills in order to efficiently decode words (e.g., monitoring letter-to-sound access) and subsequently enhance the consolidation of whole word forms in episodic memory. Hence, in this case, metacognitive monitoring skills may be of crucial role, while on the contrary, students who are more advanced in the reading developmental process, may rely more on global reading and automatic visual word recognition and have less need to monitor their performance.

From Grade 1 to Grade 3, the time points in which this study was conducted, typically developing students, based on Ehri (1998), are expected to move to the consolidated alphabetic stage and to have automatized the visual word identification process (Ehri, 1998). The fact that the lower performing students in Grade 1 were also the ones showing greater type-1 learning improvements across time in tasks assessing orthographic lexical access and visual word recognition (standardized reading accuracy), but also that the interindividual variability in the standardized reading accuracy and time reduced across time, indicates that less experienced students in reading in Grade 1 may catch up, until Grade 3, with their peers who begin primary school with fluent reading skills in place. As mentioned above, these were the students exhibiting higher metacognitive efficiency in the lexical decision task in Grade 1, which indicates an instrumental role of metacognition in regulating long-term learning improvements in the reading tasks, and assisting reading skills that are being developed (analytic reading, decoding, less automatized word identification). When these students catch up with their peers, and consolidate orthographic lexical access, metacognition seems to play a less important role. In fact, in Chapter III, we showed that early task sensitivity in the lexical decision task in lower performing students, significantly predicted decreases in their metacognitive efficiency in the same task, indicating that as they develop fluent reading between the two time points of the study, metacognitive monitoring may be less necessary..

On the contrary, students who were performing better in Grade 1 were the ones who improved less across time in type-1 reading performance and their metacognitive efficiency regarding orthographic lexical access did not appear to change across the group. It is important to mention that in the second time point of the study, these students had less room for improvement between the two time points in the standardized reading accuracy, but could still significantly improve in reading time. We propose that there may be a fine equilibrium between employing cognitive resources to efficiently and explicitly monitor one's performance in order to make essential adjustments, and performing in a perceptual type-1 task in a more automatized fashion (e.g, automatic visual word recognition), which may favour speed or the release of metacognitive resources which can be devoted on a less developed skill (e.g., reading comprehension). In this study, we provided evidence towards the reciprocal and inverse relationship of type-1 and type-2 performance by showing that increased metacognitive efficiency, mediated by decreased type-1 performance, predicted type-1 learning improvements in visual word recognition and standardized reading ability. Therefore, we suggest that explicit metacognitive monitoring may be most beneficial when a skill is developing and the student can use monitoring as a tool to inform and optimize/increase automaticity in their type-1 performance.

Conversely, when processes have already been automatized or when prerequisite skills necessary for performing in a type-1 task are very immature or not in place, explicit metacognitive monitoring may not be helpful. For instance, reflecting on single grapheme-to-phoneme correspondence may not be useful for a student who is already in the full alphabetic stage and has mastered these decodings, or vice versa, reflecting on text comprehension may not be beneficial for a student who is still learning how to map graphemes into phonemes. This view has been previously expressed by Norman (2020), who highlighted the importance of defining the settings in which promoting metacognitive thinking is beneficial and reviewed the instances in which metacognitive monitoring may hinder task performance (Norman, 2020).

We propose that researchers and practitioners would benefit by changing the focus of training metacognition in educational contexts as the panacea for improving students' learning, to primarily closely understand the context in which metacognition naturally supports learning improvements in perceptual task performance. Shedding light on this issue, will allow researchers and educationalists to optimize the design of educational interventions training metacognitive efficiency, including the optimal timing an intervention will take place, its duration, and whether students would benefit from a whole-class or individualized program training metacognition.

5.2. Outstanding questions and future directions

The present study followed a longitudinal within-subject design to examine the role of metacognitive monitoring in the first stages of reading acquisition, especially in relation to orthographic lexical processing. To our knowledge, this is the first study to track metacognitive efficiency longitudinally during development, and to assess how early metacognitive efficiency relates to the students' quantified long term learning in the context of visual word recognition.

Even though our findings provided evidence towards the beneficial role of metacognition in students' long-term improvements in visual word recognition during reading acquisition, especially in students who have not yet developed automatized reading skills, reading acquisition requires the employment of a complex set of skills, including cognitive and linguistic factors, which were not studied in the present study. Therefore, it remains unclear whether the reported findings can be generalized, for instance, to the development of students' phonological and semantic lexicon and their prerequisite skills. Further research is needed to address the role of metacognition in these skills which are equally crucial for an individual's reading development. A potential future step could be to use an identical experimental design to examine the role of metacognition on the development of phonological skills (e.g., phonemic awareness, listening skills). Moreover, it would be interesting to assess how and if training metacognition by providing feedback on participants' metacognitive judgments, would alter or enhance the role of metacognition in regulating long term learning.

Further, more research is needed to define the factors underlying the decrease in metacognitive efficiency that was observed across domains (VA span, lexical decision, emotion recognition) between Grade 1 and Grade 3, and which has been also observed in previous studies assessing metacognitive efficiency in consecutive sessions in adults (Band et al., 2019). It is important to understand whether this decrease has a functional role in the improvement of individuals' type-1 performance by potentially freeing-up brain resources devoted to reflecting on our responses, or whether the non-perceptual factors accumulated during task performance (e.g., confidence leaks across trials or experimental sessions, action fluency effects etc.) are detrimental to metacognition, and students' would benefit from explicit training enhancing metacognition in the tasks. Finally, the domain-generality/specificity issue would benefit by further investigation of the brain mechanisms employed in different tasks, and specifically, whether factors such as the age of the participants and the amount of experience participants have with the task play a role in how domain-generality/specificity of metacognition is expressed in the brain.

Overall, the present study together with other recent studies researching the role of metacognition in development, have made substantial contributions and raised important research questions related to how metacognition can support learning and academic achievement, within the fields of Reading Development, Cognitive Science, Developmental Psychology and Educational Research. Our findings contribute to this literature by emphasizing that even if metacognition is not related to students' reading ability in a certain time point of study, it can have a crucial role in students' long-term improvements in visual word recognition, working against the grain, but united, with perceptual type-1 performance. Answering the outstanding questions arising from the current dissertation, will give new insight into how we can better support in practice students who are in the process of acquiring new skills that have a great impact on human's everyday lives, such as being able to read fluently.

6. Resumen amplio en castellano

La reflexión sobre nuestros conocimientos se considera desde hace tiempo una habilidad esencial para el aprendizaje. El concepto de pensar sobre nuestras propias experiencias y nuestra capacidad de reflexionar sobre nuestra propia cognición y comportamiento, se remonta a los filósofos de la antigüedad, y sigue siendo investigado hasta la fecha bajo el término "metacognición" (Flavell, 1979; Metcalfe & Shimamura, 1994). La presente tesis pretende investigar cómo la metacognición está implicada en los primeros años de la adquisición de la lectura, que son fundamentales para el desarrollo de la lectura fluida de textos, una habilidad compleja y crucial para nuestra vida cotidiana. A continuación, ofrecemos unos breves antecedentes sobre la literatura de la metacognición y la lectura (los antecedentes ampliados se encuentran en el Capítulo I), y discutimos los principales resultados del presente estudio.

Antecedentes conceptuales

El término “metacognición” fue introducido por primera vez por Flavell (1979), quien lo clasificó en dos componentes: la metacognición declarativa, que se refiere al conocimiento declarativo de un individuo sobre su propia cognición, y la metacognición procedimental, que engloba los procesos cognitivos de orden superior que tienen lugar cuando un individuo está realizando una tarea cognitiva, y que implican la regulación de nuestro propio rendimiento en curso (Flavell, 1979).

Los modelos influyentes de metacognición procedimental proponen que la metacognición está mediada por una interacción entre un proceso a nivel de objeto (por ejemplo, una tarea de lectura o percepción, es decir, el nivel de objeto o rendimiento de tipo 1) y un proceso de segundo orden (por ejemplo, el metanivel o rendimiento de tipo 2, véase la figura 1). El componente de metanivel supervisa el proceso de primer orden y, cuando la cognición falla (es decir, tras un error), ejerce procesos de control para promover un comportamiento adaptativo (Koriat & Goldsmith, 1996; Nelson, 1990). Estos dos procesos se han denominado "supervisión metacognitiva" y "control metacognitivo", respectivamente.

Una herramienta que actualmente se utiliza ampliamente en los estudios empíricos de psicología experimental y neurociencia cognitiva (por ejemplo, Baird et al., 2013; Fleming & Lau, 2014; McCurdy et al., 2013), para evaluar la monitorización metacognitiva en el laboratorio, y que también se utilizará en el presente estudio, es la recogida de juicios de confianza retrospectivos ensayo a ensayo (respuestas de tipo 2).

Tras la decisión de primer orden (respuesta de tipo 1) relacionada con la tarea principal (por ejemplo, discriminar la categoría de secuencias de letras presentadas brevemente que representan palabras o pseudopalabras), se pide a los participantes que califiquen su grado de confianza en la corrección de la respuesta de primer orden de la tarea (Fleming & Lau, 2014). Un individuo metacognitivo ideal asigna índices de confianza altos a las respuestas correctas de tipo 1 y índices de confianza bajos a las decisiones incorrectas de tipo 1.

Los índices metacognitivos para cada individuo se estiman calculando estadísticamente el grado en que los juicios de confianza subjetivos se alinean con la precisión objetiva en el desempeño de la tarea. Esta relación se midió inicialmente utilizando diferentes índices metacognitivos (por ejemplo, correlaciones phi y gamma, curvas AUROC2), que sin embargo han sido criticados porque pueden ser confundidos por el sesgo de confianza individual, definido como la tendencia de un participante a utilizar calificaciones de confianza más altas o más bajas en una tarea cognitiva, o por el rendimiento de tipo 1 del individuo, lo que significa que los participantes que realizan mejor la tarea de tipo 1 pueden parecer erróneamente que también tienen una mejor sensibilidad metacognitiva en comparación con sus compañeros (para una revisión ver: (Fleming & Lau, 2014).

Recientemente, Maniscalco y Lau (2012) desarrollaron un modelo libre de sesgos de confianza de la Teoría de la detección de señales (TDS) para calcular la sensibilidad metacognitiva (meta-d'), que permite estimar el rendimiento de tipo 2, controlando el rendimiento de tipo 1 del participante, definido como eficiencia metacognitiva (Mratio o meta-d'/d'). Esta medida permite realizar comparaciones significativas de la eficiencia metacognitiva entre participantes o tareas (Maniscalco & Lau, 2012; para un artículo de revisión ver: Fleming & Lau, 2014). Un marco bayesiano más reciente, basado en este modelo, también ha optimizado las estimaciones de la eficiencia metacognitiva al manejar conjuntos de datos con pocos ensayos procedentes de pacientes o niños (Fleming, 2017). Por lo tanto, esta fue la medida principal utilizada en el presente estudio.

La etapa específica del desarrollo en la que los niños comienzan a proporcionar con precisión juicios verbales explícitos de monitoreo metacognitivo acerca de su desempeño en la tarea sigue siendo objeto de debate. Se ha encontrado evidencia de un procesamiento metacognitivo temprano durante el desarrollo en los bebés (Goupil & Kouider, 2016), mientras que otros estudios han informado de que la capacidad de utilizar juicios de confianza explícitos para realizar un seguimiento eficiente del rendimiento de la tarea se desarrolla alrededor de los 5 años (Destan et al., 2014; Rohwer et al., 2012). Se ha considerado que los niños comienzan a utilizar esta información para controlar su rendimiento a los 6 años (Destan et al., 2014), lo que coincide con el primer año de escolarización, el cual conlleva un cambio en las capacidades cognitivas de los estudiantes (Brod et al., 2017). No obstante, la comprensión del desarrollo de la metacognición durante los primeros años de la escuela primaria sigue siendo incompleta. En el presente estudio, nos propusimos evaluar la eficiencia metacognitiva en niños cursando los primeros años de la escuela primaria (6-9 años).

Otra cuestión crucial en la literatura de la metacognición es si el seguimiento metacognitivo se apoya en mecanismos generales o si son específicos para cada dominio cognitivo. Un modelo general de dominio predice que un individuo con una habilidad metacognitiva mala/buena en un dominio (por ejemplo, el rendimiento ortográfico), tendrá una habilidad metacognitiva mala/buena en un dominio diferente no relacionado (por ejemplo, el reconocimiento de las emociones), apoyando así la opinión de que un único sistema metacognitivo supervisa el rendimiento en diferentes dominios. Los estudios sobre el desarrollo que evalúan esta cuestión son todavía muy limitados. Estudios recientes sugieren que durante la infancia media hay un cambio gradual de los recursos específicos de dominio a los recursos generales que apoyan la metacognición (Geurten et al., 2018; Lyons & Ghetti, 2010; Vo et al., 2014). Sin embargo, aún no está claro en qué edad se produce este cambio. Investigar cómo funciona este sistema durante el primer año de adquisición de la lectura ayudará a los investigadores y educadores a comprender si la capacidad metacognitiva puede potenciarse de forma holística, es decir, en todos los dominios, o si las estrategias metacognitivas relacionadas con la adquisición de la lectura debe de asistirse por separado.

La adquisición de la lectura requiere el empleo de un conjunto complejo de habilidades cognitivas y lingüísticas (es decir, procesamiento visual de bajo nivel, procesamiento fonológico, procesos lingüísticos de alto nivel para acceder a la semántica de las palabras impresas), que se desarrollan desde una edad temprana, antes y mientras el individuo aprende a leer (Georgiou et al., 2012; González-Valenzuela et al., 2016). Existe una gran cantidad de literatura centrada en medir cómo estas habilidades influyen en la destreza lectora, en las etapas de desarrollo en las que un individuo las adquiere y en los componentes y procesos esenciales necesarios para que uno se transforme de un lector novato a un lector hábil, que decodifica y comprende eficientemente un texto impreso.

En el presente estudio, nos centramos en comprender el papel de la metacognición en el desarrollo del léxico ortográfico de los individuos, que se refiere al conocimiento ortográfico y al procesamiento de formas de palabras completas (reconocimiento visual de palabras) y es esencial para que los individuos desarrollen una lectura rápida y fluida (Ehri, 2014; Frith, 1985). El papel de la metacognición en la adquisición del conocimiento ortográfico se estudiará aquí a través de la medición de las tres habilidades siguientes.

(i) La concentración de la atención visual (CAV): la CAV es una habilidad de procesamiento pre-ortográfico que permite el procesamiento paralelo y la identificación de múltiples elementos visuales bajo la fijación de un ojo. Se ha demostrado que el procesamiento simultáneo de cadenas de letras facilita la formación de rastros ortográficos léxicos en la memoria para su acceso futuro (Bosse et al., 2007; Bosse & Valdois, 2009). La CAV se ha asociado con la velocidad de lectura en voz alta de los estudiantes de cualquier elemento de lectura y con la precisión de lectura de palabras irregulares (Bosse & Valdois, 2009; Germano et al., 2014), pero también con las habilidades de deletreo de una sola palabra durante los primeros años de la escuela primaria, tanto en ortografías superficiales como profundas (Ginestet et al., 2019; Niolaki et al., 2013).

(ii) Decisión léxica ortográfica: mide la eficiencia para encontrar una representación ortográfica de una palabra impresa en el propio léxico ortográfico y, por tanto, la calidad de esta representación (conocimiento ortográfico léxico), y se basa, en parte, en la eficiencia de la CAV (Chetail, 2017; Ginestet et al., 2019). Además, se ha comprobado que el rendimiento de los individuos en la tarea de decisión léxica predice significativamente las habilidades de identificación de palabras y también explica una varianza sustancial en la precisión de la lectura de palabras en voz alta pero no en la comprensión lectora (Katz et al., 2012).

En general, las habilidades antes mencionadas (CAV, conocimiento léxico ortográfico) se han considerado esenciales en el proceso de reconocimiento visual de palabras y para permitir el desarrollo de la lectura automática. Se ha sugerido que el aumento de la automaticidad de un individuo en el reconocimiento visual de palabras libera recursos cognitivos, que a su vez pueden emplearse en la comprensión lectora (Verhoeven et al., 2019). Por esta razón, la automaticidad en el reconocimiento de palabras se ha considerado un importante predictor del rendimiento académico en la lectura (Cunningham & Stanovich, 1997).

Estudios anteriores en poblaciones de desarrollo típico señalan el papel de la metacognición en la comprensión lectora, especialmente después de la infancia media, cuando los niños suelen desarrollar una lectura fluida. El presente estudio es, hasta donde sabemos, el primero en examinar el papel de la metacognición en el reconocimiento visual de palabras y el procesamiento ortográfico. Entender cómo y si la supervisión metacognitiva está implicada en el desarrollo de estas habilidades puede proporcionar ideas útiles sobre cómo apoyar a los niños que empiezan a leer en un contexto educativo. Con este fin, empleamos un enfoque longitudinal para evaluar la metacognición de los estudiantes en tareas relacionadas con el procesamiento léxico ortográfico y también en una tarea no relacionada con la lectura, y además valoramos el rendimiento de lectura estandarizado de los estudiantes (precisión y tiempo de lectura en listas de palabras y pseudopalabras, medido con PROLEC-R, ver detalles en Cuetos, Rodrigues, Ruano, 1996) en los primeros años de la escuela primaria (Grado 1 a Grado 3). El objetivo era proporcionar nuevos conocimientos sobre los siguientes temas:

(i) El desarrollo de la metacognición en tareas relacionadas con el procesamiento léxico ortográfico (frente a una tarea no lingüística, de reconocimiento de emociones) durante los primeros años de primaria.

(ii) El uso de mecanismos generales o específicos del dominio que apoyan la metacognición en tareas relacionadas con el procesamiento léxico ortográfico y una tarea no relacionada con la lectura durante la primera infancia.

(iii) El papel de la metacognición en la regulación del aprendizaje de los alumnos, en concreto, durante el desarrollo del léxico ortográfico y la capacidad lectora de los alumnos.

Estos objetivos de investigación se abordaron en tres capítulos (II, III y IV) con la siguiente estructura. En el capítulo II, nos centramos en las intercorrelaciones entre el rendimiento de los participantes en la tarea de tipo 1 en las tareas experimentales lingüísticas y no lingüísticas y el rendimiento de lectura estandarizado de los participantes, así como el desarrollo de estas variables a lo largo de los primeros años de la escuela primaria. Al hacerlo, sentamos las bases para explorar posteriormente el papel de la metacognición en la monitorización del rendimiento en las mismas tareas. En el capítulo III, abordamos el primer y el segundo tema mencionados anteriormente, investigando los posibles cambios en la capacidad metacognitiva a través del tiempo y probando las asociaciones en la metacognición de los estudiantes en tareas experimentales lingüísticas y no lingüísticas. Por último, en el capítulo IV, comprobamos las asociaciones entre la metacognición de los alumnos y el rendimiento en las tareas que evaluaban el procesamiento léxico ortográfico dentro y a lo largo de los primeros años de la escuela primaria. A continuación presentamos un breve resumen de los resultados de cada capítulo.

Capítulo II: Sentando las bases para explorar el papel de la metacognición en niños de primaria: Caracterización del rendimiento en tareas lingüísticas y no lingüísticas.

Como se ha mencionado anteriormente, la adquisición ortográfica está relacionada con el procesamiento eficiente de las formas ortográficas de palabras completas y desempeña un papel crucial en el desarrollo de la fluidez lectora (Adams, 1994; Stanovich, 1980). En el presente capítulo, nos propusimos evaluar la relación entre el rendimiento de las tareas de tipo 1 en las tareas lingüísticas y no lingüísticas utilizadas en la presente tesis, con el fin de sentar las bases para luego explorar el papel del monitoreo metacognitivo de las mismas tareas durante los primeros años de la escuela primaria. En concreto, investigamos: a) si la sensibilidad a la tarea en tareas relacionadas con el procesamiento léxico ortográfico (tarea del intervalo de AV, decisión léxica) y una tarea no lingüística, no relacionada con las habilidades lectoras (reconocimiento de emociones), se asocia con el rendimiento de los estudiantes en las pruebas de lectura estandarizadas (precisión lectora y tiempo de lectura) dentro y a través de los dos puntos temporales del estudio longitudinal (Grados 1 a 3), b) si el rendimiento en las mismas tareas cambia a lo largo de los dos primeros años de primaria y c) si el rendimiento de los estudiantes en el punto temporal 1 se correlaciona con su mejora de aprendizaje en cada tarea.

Nuestros resultados indicaron que el rendimiento de los alumnos en la tarea de decisión léxica, que refleja su conocimiento ortográfico, se correlacionó significativamente con su capacidad de lectura estandarizada tanto en el primer curso como en el tercero. Este resultado está en consonancia con la literatura anterior que sugiere que el conocimiento ortográfico se correlaciona con la lectura de un solo elemento y el reconocimiento de palabras (Katz et al., 2012). El rendimiento de los estudiantes en la tarea del intervalo de AV, que es un prerrequisito del procesamiento léxico ortográfico, se correlacionó significativamente o marginalmente con su capacidad de lectura estandarizada en el primer grado, pero no en el tercer grado, lo que indica que la contribución de la CAV en la lectura de un solo elemento es más fuerte en el primer año de la escuela primaria, cuando los niños reciben por primera vez la instrucción formal de la lectura (Bosse & Valdois, 2009). Como se esperaba, no se encontró ninguna asociación entre la tarea no lingüística (reconocimiento de emociones) y las variables de lectura. A continuación, se mostró una mejora significativa en todas las variables relacionadas con la lectura entre el primer y el tercer grado, siguiendo los hallazgos de estudios anteriores que sugieren que los marcadores del procesamiento léxico ortográfico y la lectura de un solo elemento muestran una mejora acelerada a lo largo de los 3 primeros años de la escuela primaria (Popa, 2020). Por último, demostramos que el rendimiento temprano en tareas relacionadas con el procesamiento léxico ortográfico (intervalo de AV, decisión léxica) predijo negativamente las mejoras longitudinales dentro de cada tarea y en tareas de lectura estandarizadas. Además, la variabilidad interindividual en las medidas de lectura disminuyó con el tiempo. Estos resultados indican que los lectores menos hábiles en el primer grado son los que mejoran más con el tiempo para alcanzar a sus compañeros más experimentados, como se ha sugerido previamente en la literatura (Suggate et al., 2013).

Capítulo III: Trayectorias de desarrollo de la metacognición en las primeras etapas de la escuela primaria: Perspectivas sobre el debate generalidad/especificidad del dominio.

Como se mencionó en los Antecedentes Conceptuales, se ha encontrado evidencia de un procesamiento metacognitivo temprano durante el desarrollo ya desde la infancia (Goupil & Kouider, 2016), mientras que otros estudios han informado que la capacidad de utilizar juicios de confianza explícitos para seguir eficientemente el rendimiento de la tarea se desarrolla más tarde durante el desarrollo alrededor de los 5 años (Destan et al., 2014; Rohwer et al., 2012). Sin embargo, la comprensión del desarrollo de la metacognición durante los primeros años de la escuela primaria sigue siendo incompleta, y existen muy pocos estudios que hagan un seguimiento longitudinal de esta capacidad durante la infancia. Además, aún se desconoce si en esta etapa de desarrollo, la metacognición se apoya en mecanismos generales o específicos para cada dominio de tarea, y si los mecanismos que favorecen la metacognición cambian durante la infancia. Por lo tanto, el presente capítulo tiene como objetivo investigar lo siguiente: a) cómo la eficiencia metacognitiva en tareas lingüísticas (CAV, tarea de decisión léxica) y tareas no lingüísticas (reconocimiento de emociones) utilizadas en la presente tesis, se desarrolla a lo largo del tiempo, y b) si se apoya en mecanismos generales o específicos del dominio a través de estas diferentes tareas, y críticamente si hay algún cambio en el desarrollo con respecto a la generalidad del dominio de la metacognición en las primeras etapas de la escuela primaria.

Contrariamente a la literatura anterior que sugiere que el seguimiento metacognitivo aumenta con el tiempo (Roebers & Spiess 2017), encontramos que la metacognición de los estudiantes, indexada como eficiencia metacognitiva (Mratio), se redujo significativamente entre el primer y el tercer grado en todas las tareas lingüísticas y no lingüísticas. Sugerimos que esta reducción puede estar relacionada con el aumento del nivel de automaticidad de los niños debido a la experiencia con estas tareas, que les permite confiar menos en su sistema metacognitivo. Este hallazgo también está en consonancia con una investigación reciente que sugiere que la disminución del ruido sensorial de tipo 1 puede estar relacionada con un aumento del ruido metacognitivo (Shekhar & Rahnev 2021). Además, los estudiantes con menor sensibilidad temprana a la tarea en el primer grado fueron los que mostraron una mayor reducción en su eficiencia metacognitiva, pero también mayores mejoras en el aprendizaje de su rendimiento de tipo

1 a lo largo del tiempo, tanto en tareas lingüísticas como no lingüísticas. Por el contrario, sus compañeros de mayor rendimiento en el primer curso mostraron una menor reducción o incluso una mejora en su eficiencia metacognitiva a lo largo del tiempo. Este hallazgo puede indicar que los estudiantes que tenían un peor rendimiento en el primer curso dependían más de sus recursos metacognitivos para ponerse a la altura de sus compañeros y liberaban estos recursos con el tiempo para dedicarlos al rendimiento de tipo 1.

Por último, se encontraron pocas pruebas de mecanismos comunes de apoyo a la metacognición en la tarea de decisión léxica y de reconocimiento de emociones en el primer curso. Sin embargo, no se encontró ninguna otra asociación en la eficiencia metacognitiva de los alumnos en ningún otro par de tareas o momento del estudio. Estos resultados están más a favor de los mecanismos específicos del dominio que apoyan la metacognición en los primeros años de la escuela primaria (Bellon et al., 2019; Geurten et al., 2018; Vo et al., 2014). La asociación única entre la tarea de decisión léxica y la de reconocimiento de emociones en el Grado 1 bien puede estar relacionada con la estructura idéntica de estas dos tareas. Por ejemplo, los estudiantes del Grado 1 pueden haber utilizado la variedad de tiempos de duración de la presentación de los estímulos, como pistas para informar sobre sus calificaciones de confianza. Se ha sugerido que la utilización estratégica de pistas se desarrolla especialmente durante los primeros años de la escuela primaria (Ackerman & Koriat, 2011; Roebers et al., 2019). Por lo tanto, en el tercer grado, los estudiantes pueden haber utilizado diferentes heurísticos en estas dos tareas durante el monitoreo metacognitivo.

En general, el presente capítulo indicó que en estas primeras etapas del desarrollo de la lectura, los estudiantes que tienen un rendimiento inferior en las tareas de tipo 1 pueden utilizar más sus recursos metacognitivos para ponerse al día con sus compañeros. Sin embargo, sería relevante para futuros estudios re-examinar la heurística utilizada en cada tarea y etapa de desarrollo para "informar" los propios juicios de confianza, y cómo esto afecta a la eficiencia de su sistema metacognitivo.

Capítulo IV: El papel de la metacognición en el desarrollo de la lectura

Además de la lectura, la capacidad metacognitiva se ha considerado fundamental para el aprendizaje en otros dominios como las matemáticas, la memoria y la percepción (Kuhn 2000; Schoenfeld 2016). Los estudios educativos sugieren que los individuos con mayores habilidades de monitoreo del desempeño tienden a ser mejores aprendices (Rawson, O'Neil & Dunlosky 2011; Metcalfe & Kornell 2007). Sin embargo, hasta la fecha, las investigaciones relacionadas con la metacognición y el aprendizaje de la lectura se centran principalmente en el aspecto de la comprensión de textos y carecen de métricas robustas de la metacognición que eviten las confusiones de los sesgos de rendimiento y confianza de tipo 1. Además, la mayoría de los estudios que evalúan la relación de la metacognición con el aprendizaje, asocian la metacognición en un punto temporal determinado con el rendimiento en la tarea en un punto temporal futuro y no examinan la cantidad de cambio en el rendimiento de los estudiantes a través de los puntos temporales.

En el presente capítulo nos propusimos evaluar:

a) la relación entre la capacidad metacognitiva de los estudiantes en tareas relacionadas con el procesamiento léxico ortográfico y el reconocimiento visual de palabras (tarea de CAV, tarea de decisión léxica) y el rendimiento de los estudiantes en pruebas de lectura estandarizadas (precisión y tiempo de lectura) en los puntos temporales del estudio, y a través de ellos, durante las primeras etapas de la adquisición de la lectura (Grado 1 y Grado 3).

b) la relación entre la eficiencia metacognitiva en la CAV, la cual es un precursor del procesamiento léxico ortográfico, y la sensibilidad a la tarea de decisión léxica, que refleja el conocimiento ortográfico dentro de los dos puntos temporales del estudio longitudinal (Grados 1, 3).

c) el valor predictivo de la eficiencia metacognitiva de tipo 2 en las tareas experimentales relacionadas con la lectura al inicio del estudio sobre las mejoras en el rendimiento de los estudiantes en las tareas experimentales de tipo 1 y en la capacidad de lectura estandarizada a lo largo de los puntos temporales (Grados 1 a 3).

En general, utilizando estadísticas frecuentistas, no se encontraron asociaciones entre la eficiencia metacognitiva en las tareas que evalúan el procesamiento léxico ortográfico (intervalo de AV, decisión léxica) y el rendimiento de lectura estandarizado de los estudiantes (precisión y tiempo de lectura) ni en el Grado 1, ni en el Grado 3. Sin embargo, el análisis bayesiano proporcionó evidencia hacia una relación negativa entre: a) la eficiencia metacognitiva de los participantes en la tarea de la CAV y el rendimiento en la misma tarea y b) la eficiencia metacognitiva de los participantes en la tarea de decisión léxica en el primer grado y el rendimiento de los estudiantes en la misma tarea, en la tarea del intervalo de AV y en la precisión de lectura estandarizada en el mismo momento del estudio.

Además, los estudiantes con mayor eficiencia metacognitiva temprana en la tarea de decisión léxica en el primer grado fueron los que mostraron las mayores mejoras de aprendizaje en su rendimiento de tipo 1 a través del tiempo, tanto en esta tarea como en la precisión de lectura estandarizada. En el capítulo II, mostramos que las mayores mejoras de aprendizaje en las tareas relacionadas con la lectura correspondieron a los estudiantes que mostraron la menor sensibilidad temprana a la tarea. También demostramos que el análisis de mediación indicó que el efecto de la eficiencia metacognitiva temprana sobre las mejoras de aprendizaje en estas tareas, está mediado por la sensibilidad temprana a la tarea.

Tomando en conjunto estos dos hallazgos, sugerimos que los estudiantes que comienzan el primer grado con mayores habilidades de lectura, y han dominado sus mapeos ortográficos, pueden haber automatizado sus habilidades de lectura de palabras visuales (Verhoeven & Perfetti, 2017), y por lo tanto, tienen menos necesidad de monitorear su desempeño durante la lectura de palabras automática. Por el contrario, los estudiantes que tienen menos fluidez en la lectura en el primer año de la escuela primaria, pueden confiar más en los recursos metacognitivos, con el fin de monitorear cuidadosamente su desempeño y construir su léxico ortográfico a través de la lectura analítica.

Conclusión general

Comprender el papel de la metacognición en el control del rendimiento cognitivo y conductual y en la orientación del aprendizaje, incluida la adquisición de la lectura, durante la infancia, puede tener importantes implicaciones en el diseño de programas educativos para el entrenamiento de la metacognición y la promoción de habilidades transferibles a través de distintos dominios de rendimiento, así como para la promoción del aprendizaje permanente y la superación personal.

En la presente tesis, aportamos evidencias de que los recursos metacognitivos pueden ser utilizados como una herramienta para mejorar el rendimiento de los estudiantes de forma longitudinal y que la necesidad del empleo de estos recursos puede estar relacionada con el nivel de automaticidad alcanzado en las tareas de tipo 1. Por ejemplo, encontramos que los estudiantes que muestran un menor rendimiento de tipo 1 en las tareas que evalúan el procesamiento léxico ortográfico, pero una mayor eficiencia metacognitiva durante las primeras etapas de la adquisición de la lectura fueron los que mostraron la mayor cantidad de mejora en el rendimiento de tipo 1 a través del tiempo. Además, en el primer grado, la metacognición de los estudiantes en la tarea de decisión léxica, que refleja el conocimiento léxico ortográfico de los estudiantes, se correlacionó negativamente con CAV. Este hallazgo proporciona un apoyo adicional, que los estudiantes con una mayor CAV, que contribuye a la lectura automática de palabras a la vista, pueden tener menos necesidad de monitorear su desempeño durante el reconocimiento de palabras visuales, que sus compañeros con una menor CAV en el primer año de la escuela primaria. La relación entre la metacognición y el nivel de automatismo en las tareas de tipo 1 se demostró cuando se realizó un seguimiento longitudinal del desarrollo de la eficacia metacognitiva durante los primeros años de la escuela primaria. Tanto en las tareas lingüísticas como en las no lingüísticas, la eficiencia metacognitiva disminuyó con el tiempo, a medida que aumentaba la experiencia de los alumnos en la tarea.

Los resultados mencionados subrayan la importancia de evaluar paralelamente la eficiencia metacognitiva y el rendimiento en la tarea en las primeras etapas de la adquisición de la lectura, ya que ambas variables contribuyen al aprendizaje de la lectura a largo plazo de los alumnos. Es importante destacar que puede ser más crucial potenciar la metacognición en contextos educativos cuando un proceso (por ejemplo, el reconocimiento de palabras automático) aún no se ha automatizado o adquirido. En este último caso, liberar recursos cognitivos dedicados a la supervisión metacognitiva para aumentar el nivel de automaticidad en el procesamiento de señales de tipo 1 puede ser más beneficioso. Es necesario seguir investigando para entender si los programas que mejoran la capacidad metacognitiva deben abordar la metacognición como una habilidad transferible a través de distintos dominios, o si los dominios únicos deben dirigirse por separado y si es importante hacerlo durante el desarrollo temprano de una habilidad.

7. Related publications and manuscripts

Taouki I., Lallier M., Soto D. (accepted). “The role of metacognition in monitoring performance and regulating learning in early readers” *Metacognition and Learning*.

Taouki I., Lallier M., Soto D., Hauk O. (in preparation). “Theta oscillations support metacognitive monitoring in early childhood”.

8. References

Ackerman, R., & Koriat, A. (2011). Response latency as a predictor of the accuracy of children's reports. *Journal of Experimental Psychology. Applied*, *17*(4), 406–417.

Adams, M. J. (1994). *Beginning to Read: Thinking and Learning about Print*. MIT Press.

Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition*, *146*, 377–386.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear Probability, Logit, and Probit Models*. SAGE.

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *eLife*, *5, e18103*. https://doi.org/10.7554/eLife.18103

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An Introduction to Bayesian Hypothesis Testing for Management Research. *Journal of Management*, *41*(2), 521–543.

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, *105*(4), 678–723.

Anthony, J. L., & Francis, D. J. (2005). Development of Phonological Awareness. *Current Directions in Psychological Science*, *14*(5), 255–259.

August, D. L., Flavell, J. H., & Clift, R. (1984). Comparison of Comprehension Monitoring of Skilled and Less Skilled Readers. *Reading Research Quarterly*, *20*(1), 39–53.

Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(42), 16657–16665.

Bang, J. W., Shekhar, M., & Rahnev, D. (2019). Sensory noise increases metacognitive efficiency. *Journal of Experimental Psychology. General*, *148*(3), 437–452.

Bellon, E., Fias, W., & De Smedt, B. (2019). More than number sense: The additional role of executive functions and metacognition in arithmetic. *Journal of Experimental Child Psychology*, *182*, 38–60.

Besner, D., Risko, E. F., Stolz, J. A., White, D., Reynolds, M., O'Malley, S., & Robidoux, S. (2016). Varieties of Attention: Their Roles in Visual Word Identification. *Current Directions in Psychological Science*, *25*(3), 162–168.

Bosse, M.-L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, 104(2):198-230. https://doi.org/10.1016/j.cognition.2006.05.009

Bosse, M.-L., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: a cross-sectional study. *Journal of Research in Reading*, *32*(2), 230–253.

Brod, G., Bunge, S. A., & Shing, Y. L. (2017). Does One Year of Schooling Improve Children's Cognitive Control and Alter Associated Brain Activation? *Psychological Science*, 28(7), 967–978. https://doi.org/10.1177/0956797617699838

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, *19*(1), 5–51.

Chen, C., Schneps, M. H., Masyn, K. E., & Thomson, J. M. (2016). The Effects of Visual Attention Span and Phonological Decoding in Reading Comprehension in Dyslexia: A Path Analysis. *Dyslexia* , *22*(4), 322–344.

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, *163*, 103–120.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, *4*(1), 5–17.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(6), 1197–1211.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Coltheart, M., Saunders, S. J., & Tree, J. J. (2010). Computational modelling of the effects of semantic dementia on visual word recognition. *Cognitive Neuropsychology*, *27*(2), 101–114.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934–945.

de Bruin, A. B. H., Thiede, K. W., Camp, G., & Redford, J. (2011). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology*, *109*(3), 294–310.

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. *Psychological Science*, *29*(5), 761–778.

Destan, N., Hembacher, E., Ghetti, S., & Roebers, C. M. (2014). Early metacognitive abilities: the interplay of monitoring and control processes in 5- to 7-year-old children. *Journal of Experimental Child Psychology*, *126*, 213–228.

De Vos, T. (1992). Tempo Test Rekenen Nijmegen. *The Netherlands: Berkhout*.

Diedenhofen, B., & Musch, J. (2015). cocor: a comprehensive solution for the statistical comparison of correlations. *PloS One*, *10*(3), e0121945.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: one-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*(4), 1246–1258.

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.

Dunlosky, J., Mueller, M. L., Morehead, K., Tauber, S. K., Thiede, K. W., & Metcalfe, J. (2021). *Why Does Excellent Monitoring Accuracy Not Always Produce Gains In Memory Performance? Zeitschrift für Psychologie, 229*(2), 104–119

Dunn, O. J., & Clark, V. (1969). Correlation Coefficients Measured on the Same Individuals. *Journal of the American Statistical Association*, *64*(325), 366–377.

Efklides, A. (2008). Metacognition: Defining Its Facets and Levels of Functioning in Relation to Self-Regulation and Co-regulation. *European Psychologist*, *13*(4), 277–287.

Efklides, A. (2011). Interactions of Metacognition With Motivation and Affect in Self-Regulated Learning: The MASRL Model. *Educational Psychologist*, 46(1), 6–25. https://doi.org/10.1080/00461520.2011.538645

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, 18(2),116–125. https://doi.org/10.1111/j.1467-9817.1995.tb00077.x

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. *Word Recognition in Beginning Literacy*, *1*, 3–40.

Ehri, L. C. (2005). Learning to Read Words: Theory, Findings, and Issues. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, *9*(2), 167–188.

Ehri, L. C. (2014). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling Memory, and Vocabulary Learning. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, *18*(1), 5–21.

Fandakova, Y., Selmeczy, D., Leckey, S., Grimm, K. J., Wendelken, C., Bunge, S. A., & Ghetti, S. (2017). Changes in ventromedial prefrontal and insular cortex support the development of metamemory from childhood into adolescence. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(29), 7582–7587.

Filevich, E., Forlim, C. G., Fehrman, C., Forster, C., Paulus, M., Shing, Y. L., & Kühn, S. (2020). I know that I know nothing: Cortical thickness and functional connectivity underlying meta-ignorance ability in preschoolers. *Developmental Cognitive Neuroscience*, *41*, 100738.

Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction*, *32*, 1–9.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *The American Psychologist*, *34*(10), 906–911.

Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1), nix007.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1338–1349.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443.

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psychological Science*, *26*(1), 89–98.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating Introspective Accuracy to Individual Differences in Brain Structure. *Science*, 329(5998), 1541–1543. https://doi.org/10.1126/science.1191883

Fleur, D. S., Bredeweg, B., & van den Bos, W. (2021). Metacognition: ideas and insights from neuro- and educational sciences. *NPJ Science of Learning*, *6*(1), 13.

Frith, U. (1985). Beneath the surface of developmental dyslexia. *Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading*. https://ci.nii.ac.jp/naid/10022405906/

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*(4), 843–876.

Garner, R. (1987). Metacognition and reading comprehension. *Cognition and Literacy*, *165*. https://psycnet.apa.org/fulltext/1987-97150-000.pdf

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing*, *25*(2), 321–346.

Germano, G. D., Reilhac, C., Capellini, S. A., & Valdois, S. (2014). The phonological and visual basis of developmental dyslexia in Brazilian Portuguese reading children. *Frontiers in Psychology*, *5*, 1169.

Geurten, M., Meulemans, T., & Lemaire, P. (2018). From domain-specific to domain-general? The developmental path of metacognition for strategy selection. In *Cognitive Development*, 48, 62–81. https://doi.org/10.1016/j.cogdev.2018.08.002

Ghetti, S., Hembacher, E., & Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: A metacognitive approach. *Child Development Perspectives*, *7*(3), 160–165.

Ginestet, E., Phénix, T., Diard, J., & Valdois, S. (2019). Modeling the length effect for words in lexical decision: The role of visual attention. *Vision Research*, *159*, 10–20.

Ginestet, E., Shadbolt, J., Tucker, R., Bosse, M., & Hélène Deacon, S. (2021). Orthographic learning and transfer of complex words: insights from eye tracking during reading and learning tasks. *Journal of Research in Reading*, 44(1), 51–69. https://doi.org/10.1111/1467-9817.12341

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology. Human Perception and Performance*, *5*(4), 674–691.

González-Valenzuela, M. J., Díaz-Giráldez, F., & López-Montiel, M. D. (2016). Cognitive Predictors of Word and Pseudoword Reading in Spanish First-Grade Children. *Frontiers in Psychology*, *7*, 774.

Goupil, L., & Kouider, S. (2016). Behavioral and Neural Indices of Metacognitive Sensitivity in Preverbal Infants. *Current Biology: CB*, *26*(22), 3038–3045.

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. In *Proceedings of the National Academy of Sciences* (Vol. 113, Issue 13, pp. 3492–3496). https://doi.org/10.1073/pnas.1515129113

Green, D. M., & Swets, J. A. (1988). *Signal Detection Theory and Psychophysics*. Peninsula Pub.

Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, *37*(7), 1001–1013.

Hains, S. M. J. (1986). The Role of Word Familiarity in Word Recognition. In S. E. Newstead, S. H. Irvine, & P. L. Dann (Eds.), *Human Assessment: Cognition and Motivation* (pp. 179–199). Springer Netherlands.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological Review*, *106*(3), 491–528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720.

Holmes, V. M., & Dawson, G. (2014). Visual-attentional span and lexical decision in skilled adult readers. *Journal of Research in Reading*, *37*(4), 331–355.

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The Complex Nature of Reading Fluency: A Multidimensional View. In *Reading & Writing Quarterly* (Vol. 25, Issue 1, pp. 4–32). https://doi.org/10.1080/10573560802491208

Jacobs, J. E., & Paris, S. G. (1987). Children's Metacognition About Reading: issues in Definition, Measurement, and Instruction. *Educational Psychologist*, *22*(3-4), 255–278.

Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J., & Whalen, D. H. (2012). What lexical decision and naming tell us about reading. *Reading and Writing*, *25*(6), 1259–1282.

Kelly, M., Scholnick, E. K., Travers, S. H., & Johnson, J. W. (1976). Relations among memory, memory appraisal, and memory strategies. *Child Development, 47, 648-659*.

Koriat, A., & Ackerman, R. (2010). Metacognition and mindreading: judgments of learning for Self and Other during self-paced study. *Consciousness and Cognition*, *19*(1), 251–264.

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The easily learned, easily remembered heuristic in children. *Cognitive Development*, *24*(2), 169–182.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517.

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*, *18*(1), 64–71.

Krebs, S. S., & Roebers, C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *The British Journal of Educational Psychology*, *80*(Pt 3), 325–340.

Kuhn, D. (2000). Metacognitive Development. *Current Directions in Psychological Science,* 9(5), 178–181. https://doi.org/10.1111/1467-8721.00088

Kwon, M., Legge, G. E., & Dubbels, B. R. (2007). Developmental changes in the visual span for reading. *Vision Research*, *47*(22), 2889–2900.

Lallier, M., Acha, J., & Carreiras, M. (2016). Cross-linguistic interactions influence reading development in bilinguals: a comparison between early balanced French-Basque and Spanish-Basque bilingual children. *Developmental Science*, *19*(1), 76–89.

Lallier, M., Donnadieu, S., & Valdois, S. (2013). Developmental dyslexia: exploring how much phonological and visual attention span disorders are linked to simultaneous auditory processing deficits. *Annals of Dyslexia*, *63*(2), 97–116.

Lawrence, K., Campbell, R., & Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Frontiers in Psychology*, *6*, 761.

Lecce, S., Caputi, M., & Pagnin, A. (2015). False-belief understanding at age 5 predicts beliefs about learning in year 3 of primary school. *The European Journal of Developmental Psychology*, *12*(1), 40–53.

Lipko, A. R., Dunlosky, J., Lipowski, S. L., & Merriman, W. E. (2012). Young children are not underconfident with practice: The benefit of ignoring a fallible memory heuristic. *Journal of Cognition and Development: Official Journal of the Cognitive Development Society*, *13*(2), 174–188.

Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Developmental Psychology*, *49*(8), 1505–1516.

Lockl, K., & Schneider, W. (2006). Precursors of metamemory in young children: the role of theory of mind and metacognitive vocabulary. *Metacognition and Learning*, 1(1), 15–31. https://doi.org/10.1007/s11409-006-6585-9

Lockl, K., & Schneider, W. (2007). Knowledge about the mind: links between theory of mind and later metamemory. *Child Development*, *78*(1), 148–167.

Lyons, K. E., & Ghetti, S. (2010). Metacognitive Development in Early Childhood: New Questions about Old Assumptions. *Trends and Prospects in Metacognition Research*, 259–278. https://doi.org/10.1007/978-1-4419-6546-2_12

Maniscalco, B., & Lau, H. (2011). On a distinction between detection and discrimination: metacognitive advantage for signal over noise. *Journal of Vision,* 11(11), 163–163. https://doi.org/10.1167/11.11.163

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Maniscalco, B., McCurdy, L. Y., Odegaard, B., & Lau, H. (2017). Limited Cognitive Resources Explain a Trade-Off between Perceptual and Metacognitive Vigilance. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(5), 1213–1224.

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: implications for studies of metacognitive processes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(2), 509–527.

Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. Retrospective confidence judgments across tasks: domain-general processes underlying metacognitive accuracy. *PsyArXiv.* [Preprint.] June 18 2020 [accessed 2021 September] https://doi.org/10.31234/osf.io/dr7ba

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(5), 1897–1906.

Metcalfe, J. (2009). Metacognitive Judgments and Control of Study. *Current Directions in Psychological Science*, *18*(3), 159–163.

Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning*, *8*(1), 19–46.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: the effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*(2), 225–229.

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about Knowing*. MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001

Meuwissen, A. S., Anderson, J. E., & Zelazo, P. D. (2017). The creation and validation of the Developmental Emotional Faces Stimulus Set. *Behavior Research Methods*, *49*(3), 960–966.

Moelands, F., & Rymenans, R. (2003). Schaal vorderingen in spellingvaardigheid voor vlaanderen (SVS-V). Handleiding. In *Flemish scale for spelling skills (SVS-V). Manual. Arnhem*. The Netherlands: Citogroep.

Molenberghs, P., Trautwein, F.-M., Böckler, A., Singer, T., & Kanske, P. (2016). Neural correlates of metacognitive ability and of feeling confident: a large-scale fMRI study. *Social Cognitive and Affective Neuroscience*, *11*(12), 1942–1951.

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(14), 3534–3546.

Moses-Payne, M. E., Habicht, J., Bowler, A., Steinbeis, N., & Hauser, T. U. (2021). I know better! Emerging metacognition allows adolescents to ignore false advice. *Developmental Science*, *24*(5), e13101.

Murphy, M. M., & Mazzocco, M. M. M. (2008). Mathematics learning disabilities in girls with fragile X or Turner syndrome during late elementary school.*Journal of Learning Disabilities*, *41*(1), 29–46.

Narens, T. O. (1990). Metamemory: A Theoretical Framework and New Findings. *Psychology of Learning and Motivation*, 26, 125–173. https://doi.org/10.1016/s0079-7421(08)60053-5

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of Metacognitive Judgments in the Allocation of Study During Multitrial Learning. *Psychological Science,* 5(4), 207–213. https://doi.org/10.1111/j.1467-9280.1994.tb00502.x

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. *Metacognition: Knowing about Knowing*, *13*, 1–25.

Niolaki, G. Z., Masterson, J., & Terzopoulos, A. R. (2013). Association of single word spelling with visual attention span and phonological ability in monolingual English-and Greek-speaking children. *13th European Congress of Psychology, Stockholm, Sweden*.

Norman, E. (2020). Why Metacognition Is Not Always Helpful. *Frontiers in Psychology*, *11*, 1537. https://doi.org/10.3389/fpsyg.2020.01537

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of Comprehension Monitoring and Working Memory in Good and Poor Comprehenders. *Reading and Writing*, *18*(7), 657–686.

Otero, J., & Kintsch, W. (1992). Failures to Detect Contradictions in a Text: What Readers Believe versus what they Read. *Psychological Science*, *3*(4), 229–236.

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, *28*, 151–160.

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: a review. *Neuropsychology Review*, *15*(3), 105–130.

Pearson, K., & Filon, L. N. G. (1897). Mathematical Contributions to the Theory of Evolution. IV. On the Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation. *Proceedings of the Royal Society of London*, *62*, 173–176.

Peirce J & Macaskill. (2019). Building Experiments in PsychoPy, *Perception*, 48(2), 189–190. https://doi.org/10.1177/0301006618823976

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115.

Popa, R. M. (2020). PROLEC BATTERY-Screening Tool for Dyslexia. *Revista Română de Terapia Tulburărilor de Limbaj şi Comunicare,* 4(2), 24–33. https://doi.org/10.26744/rrttlc.2020.6.2.03

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891.

Preacher, K. J., & Leonardelli, G. J. (2001). *Calculation for the Sobel test: An interactive calculation tool for mediation tests*.

Raghunathan, T. E., Rosenthal, R., & Rubin, D. B. (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods*, *1*(2), 178–183.

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence Leak in Perceptual Decision Making. *Psychological Science*, *26*(11), 1664–1680.

Rahnev, D., Desender, K., Lee, A.L.F. et al. (2020). The Confidence Database. *Nature Human Behaviour*, *4*, 317–325.

Rawson, K. A., O'Neil, R., & Dunlosky, J. (2011). Accurate monitoring leads to effective control and greater learning of patient education materials. *Journal of Experimental Psychology. Applied*, *17*(3), 288–302.

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(6), 787–799.

Rinne, L. F., & Mazzocco, M. M. M. (2014). Knowing right from wrong in mental arithmetic judgments: calibration of confidence predicts the development of accuracy. *PloS One*, *9*(7), e98663.

Roebers, C. M., Mayer, B., Steiner, M., Bayard, N. S., & van Loon, M. H. (2019). The role of children's metacognitive experiences for cue utilization and monitoring accuracy: A longitudinal study. *Developmental Psychology*, *55*(10), 2077–2089.

Roebers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *The British Journal of Educational Psychology*, *79*(Pt 4), 749–767.

Roebers, C. M., & Spiess, M. (2017). The Development of Metacognitive Monitoring and Control in Second Graders: A Short-Term Longitudinal Study. *Journal of Cognition and Development: Official Journal of the Cognitive Development Society*, *18*(1), 110–128.

Rohwer, M., Kloo, D., & Perner, J. (2012). Escape from meta ignorance: how children develop an understanding of their own lack of knowledge. *Child Development*, *83*(6), 1869–1883.

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human metacognition across domains: insights from individual differences and neuroimaging. *Personality Neuroscience*, *1(e17), 1-13.* https://doi.org/10.1017/pen.2018.16

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, *1*(3), 165–175.

Ruby, E., Giles, N., & Lau, H. (2017). Finding domain-general metacognitive mechanisms requires using appropriate tasks.*BiorXiv, 1-20,* https://doi.org/10.1101/211805

Ryals, A. J., Rogers, L. M., Gross, E. Z., Polnaszek, K. L., & Voss, J. L. (2016). Associative Recognition Memory Awareness Improved by Theta-Burst Stimulation of Frontopolar Cortex. *Cerebral Cortex* , *26*(3), 1200–1210.

Samaha, J., & Postle, B. R. (2017). Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proceedings. Biological Sciences / The Royal Society*, *284*(1867). https://doi.org/10.1098/rspb.2017.2035

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*(2), 107–121.

Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, *1*(2), 291–297.

Schneider, W. (2015). *Memory Development from Early Childhood Through Emerging Adulthood*. Springer, Cham.

Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task. *Cognitive Development*, *15*(2), 115–134.

Schoenfeld, A. H. (2016). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics (Reprint). *Journal of Education*, 196(2), 1–38. https://doi.org/10.1177/002205741619600202

Schraw, G., Dunkle, M. E., & Bendixen, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology*, 9(6), 523–538.

https://doi.org/10.1002/acp.2350090605

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, 55(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129.

Share, D. L., & Shalev, C. (2004). Self-teaching in normal and disabled readers. *Reading and Writing*, 17(8), 769–800. https://doi.org/10.1007/s11145-004-2658-9

Shekhar, M., & Rahnev, D. (2018). Distinguishing the Roles of Dorsolateral and Anterior PFC in Visual Metacognition. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(22), 5078–5087.

Shekhar, M., & Rahnev, D. (2021). Sources of Metacognitive Inefficiency. *Trends in Cognitive Sciences*, *25*(1), 12–23.

Silver, N. C., Hittner, J. B., & May, K. (2004). Testing Dependent Correlations With Non overlapping Variables: A Monte Carlo Simulation. *Journal of Experimental Education*, *73*(1), 53–69.

Sodian, B., Thoermer, C., Kristen, S., & Perst, H. (2012). Metacognition in infants and young children. *Foundations of Metacognition*, 119–133. https://doi.org/10.1093/acprof:oso/9780199646739.003.0008

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(1), 204–221.

Stanovich, K. E. (1980). Toward an Interactive-Compensatory Model of Individual Differences in the Development of Reading Fluency. *Reading Research Quarterly*, *16*(1), 32–71.

Stanovich, K. E. (1993). The Language Code: Issues in Word Recognition. In S. R. Yussen & M. C. Smith (Eds.), *Reading Across the Life Span*, 111–135. Springer New York.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251.

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. *Handbook of Metacognition in Education*, 85–106.

Valdois, S., Bosse, M.-L., Ans, B., Carbonnel, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case studies. *Reading and Writing*, *16*(6), 541–572.

Valdois, S., Roulin, J.-L., & Line Bosse, M. (2019). Visual attention modulates reading acquisition. *Vision Research*, *165*, 152–161.

van den Boer, M., de Jong, P. F., & Haentjens-van Meeteren, M. M. (2013). Modeling the Length Effect: Specifying the Relation With Visual and Phonological Correlates of Reading. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, *17*(4), 243–256.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2020). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-020-01798-5

Verhoeven, L., & Perfetti, C. (2017). *Learning to Read across Languages and Writing Systems*. Cambridge University Press.

Verhoeven, L., Reitsma, P., & Siegel, L. S. (2011). Cognitive and linguistic factors in reading acquisition. *Reading and Writing*, *24*(4), 387–394.

Verhoeven, L., Voeten, M., & Vermeer, A. (2019). Beyond the simple view of early first and second language reading: The impact of lexical quality. *Journal of Neurolinguistics*, *50*, 28–36.

Vo, V. A., Li, R., Kornell, N., Pouget, A., & Cantlon, J. F. (2014). Young children bet on their numerical skills: metacognition in the numerical domain. *Psychological Science*, *25*(9), 1712–1721.

Vuorre, M., & Metcalfe, J. (2021). Measures of relative metacognitive accuracy are confounded with task performance in tasks that permit guessing. *Metacognition and Learning, 16, 1-23.* https://link.springer.com/article/10.1007/s11409-020-09257-1

Wechsler, D. (2014). Wechsler intelligence scale for children–Fifth Edition (WISC-V). *Bloomington, MN: Pearson*.

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., Dolan, R. J., & Blakemore, S.-J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, *22*(1), 264–271.

Wendelken, C., Ferrer, E., Ghetti, S., Bailey, S. K., Cutting, L., & Bunge, S. A. (2017). Frontoparietal Structural Connectivity in Childhood Predicts Development of Functional Connectivity and Reasoning Ability: A Large-Scale Longitudinal Investigation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(35), 8549–8558.

Wickens, T. D. (2001). *Elementary Signal Detection Theory*. https://doi.org/10.1093/acprof:oso/9780195092509.001.0001

Winograd, P., & Johnston, P. (1982). Comprehension Monitoring and the Error Detection Paradigm. *Journal of Reading Behavior*, *14*(1), 61–76.

Xue, K., Shekhar, M., & Rahnev, D. (2021). Examining the robustness of the relationship between metacognitive efficiency and metacognitive bias. *Consciousness and Cognition*, *95*, 103196.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1310–1321.

Zarić, J., Hasselhorn, M., & Nagler, T. (2021). Orthographic knowledge predicts reading and spelling skills over and above general intelligence and phonological awareness. *European Journal of Psychology of Education*, *36*(1), 21–43.

Ziegler, J. C., & Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin*, 131(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. *Handbook of metacognition in education*, 311–328. Routledge.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*(4), 399–413.

9. Appendix A

9.1. A1. Information Sheet for the legally responsible of the minors

Date: …../…./…..

Title of the project: The role of metacognition in reading acquisition

Investigators: Ioanna Taouki, Predoctoral Researcher BCBL
Marie Lallier, Staff Scientist BCBL
David Soto, Research Professor BCBL

Invitation to take part in a research study

Your child is invited to take part in a research study. Before you decide about whether you are happy for them to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and, if you wish, discuss it with friends and relatives. Ask us if there is anything that is unclear or if you would like more information. Take time to decide whether or not you would like your child to participate in the study. Thank you for reading this.

Information about the study

The study aims at understanding how learning to read is related to the development of the so-called metacognitive skills. Metacognition means "cognition about cognition", "thinking about thinking", "knowing about knowing", in other words being aware of what I am learning and what I already know. We consider it important that teachers help students develop the ability to reflect critically on learning experiences, which will inform their future progress and make them feel more confident about what they already know, but also be aware of what they need to improve. With this study we anticipate to shed light on the role of metacognition in reading acquisition, a learning process that is complex and builds on cognitive, linguistic, and social skills developed from a very early age.

We are hoping to understand how metacognition naturally develops as an ability through the first years of primary school that students learn how to read. That will help us discuss the importance of enhancing meta-cognitive skills in classrooms from an early age.

What does the study involve?

The study involves two identical experimental parts. In Part 1, at the beginning of the year we will give your child short reading and language tests in Spanish such as reading words, or strings of letters, or recognizing shapes presented in a computer screen. Each test will include the evaluation of the metacognitive ability of your child asking them to rate their confidence on how well they did in each task. In Part 2, the same tests are going to be repeated, after children have received 2 years of formal instruction in reading. Both experimental parts (Part 1 and 2) will take place in your child's school, during school time, and according to a schedule set up in accordance with your child's teacher and director. The total duration of each part will be 6 sessions of 30 minutes. You or your child will not be paid for your participation in the study. Your child will receive a certificate at the end of the study for participating in this research.

Are there any benefits or risks?

This study has been reviewed by the Research Ethics Committee of the University of Barcelona and the Research Committee of the Basque Center on Cognition, Brain and Language and there is absolutely no risk of harming your child, as we are going to use only behavioural measurements in a computer screen. In the unlikely case of distress or uncomfortable feeling experienced by your child, teachers and experimenters will be there to ensure that the experimental session will stop at any time for the welfare of your child.

What will happen to my data?

All data collected will be confidential, and you or your child will not be identifiable in any report, thesis or publication which arises from this study. The data from this study will be stored securely during the project and destroyed at the end. Because this experiment is not directed to diagnostic or clinical purposes but research purposes only, we cannot give you any direct feedback concerning how your child performed or his/her individual results. However, we will be happy to provide a summary of our *group* findings when the study is complete. If you or your child choose to withdraw from the study and your child's data is identifiable to the research team, then you have the right to request that your child's data is not used. There will be no commercial exploitation of this research.

What if I/my child don't want to take part?

It is up to you to decide whether or not your child would like to participate in this study. Deciding not to take part will not impact any other aspect of your professional activity or your child's school activity. If you do decide that your child can take part, you will be given this information sheet to keep and will be asked to sign a consent form. You will be free to withdraw at any time and without giving a specific reason.

Who do I contact about the study?

If you have further questions regarding this study, please feel free to contact:

Ioanna Taouki, predoctoral researcher, BCBL

e-mail: i.taouki@bcbl.eu, Tel: 633308117

Who do I contact with any concerns about this study?

In the event of any complaints arising concerning this research, please address them to info@bcbl.eu.

Thank you for considering your participation in this study!
Ioanna Taouki, Predoctoral Researcher

PLEASE KEEP THIS INFORMATION SHEET AND A COPY OF SIGNED CONSENT FORM FOR YOUR RECORD.

9.2. A2. Consent Form for the legally responsible of the minors

Title of the project: The role of metacognition in reading acquisition
Date:
Participant code:
Investigators: Ioanna Taouki, Predoctoral Researcher BCBL
Marie Lallier, Staff Scientist of BCBL
David Soto, Professor Researcher of BCBL

The investigation and my part in the investigation have been fully explained to me by one of the investigators listed above and I understand his/her explanation. I understand that my child will have to take part in visual computer-based tasks assessing his/her reading skills, and that this procedure is completely safe, as it includes only behavioural measurements. The procedures of this investigation and their risks have been answered to my satisfaction.

I understand that I/my child am/is free not to answer specific items or questions in preliminary interviews or questionnaires.

I understand that all data will be stored, analyzed and published in a completely confidential manner with regard to my identity and the identity of my child, and that me or my child is free to withdraw my or his/her consent and terminate my participation at any time without penalty.

I understand that this research is only for scientific and not diagnostic purposes, that my questions will be answered and that I may request a summary of the results of this study concerning the entire group of children tested. I understand that my participation in this research is voluntary and that I will not receive any financial compensation.
Following this, I, ..., hereby agree that my child will participate in the present scientific investigation under the supervision of the Predoctoral Researcher, Ioanna Taouki, including visual computer-based tasks assessing reading and meta-cognitive skills of my child.

I know of no medical condition (e.g., neurological impairment), which may cause adverse effects to my child if he/she participates in this research.

Signed ________________________________ Date _________________

9.3. A3. Consent Sheet for the minors

Invitation to participate in a research study !

Hi !
We invite you to help a team of researchers to find out how children learn to read, what strategies they use and which is the best way we teachers can help them learn how to read.

What will you need to do ?

Before you decide if you want to participate, it is important to know how you can help us and what we will ask you to do. This research will be done in two parts. At the beginning of the year, we are going to meet during 6 different days for half an hour. This will happen in your school and we will agree with you teacher which is the best time to do that. What we will ask you during this time is to do some small tasks like reading some words or playing little games on a computer, like playing with different shapes on the screen. At the end of the year, we are going to meet again and we will ask you to do the same tasks to see how much you have improved !
Don't worry at all about making mistakes in the tasks, nobody will be able to see your results except the team of researchers! So the most important thing is to feel comfortable and enjoy the games!
At the end of the study you will get a certificate for helping us

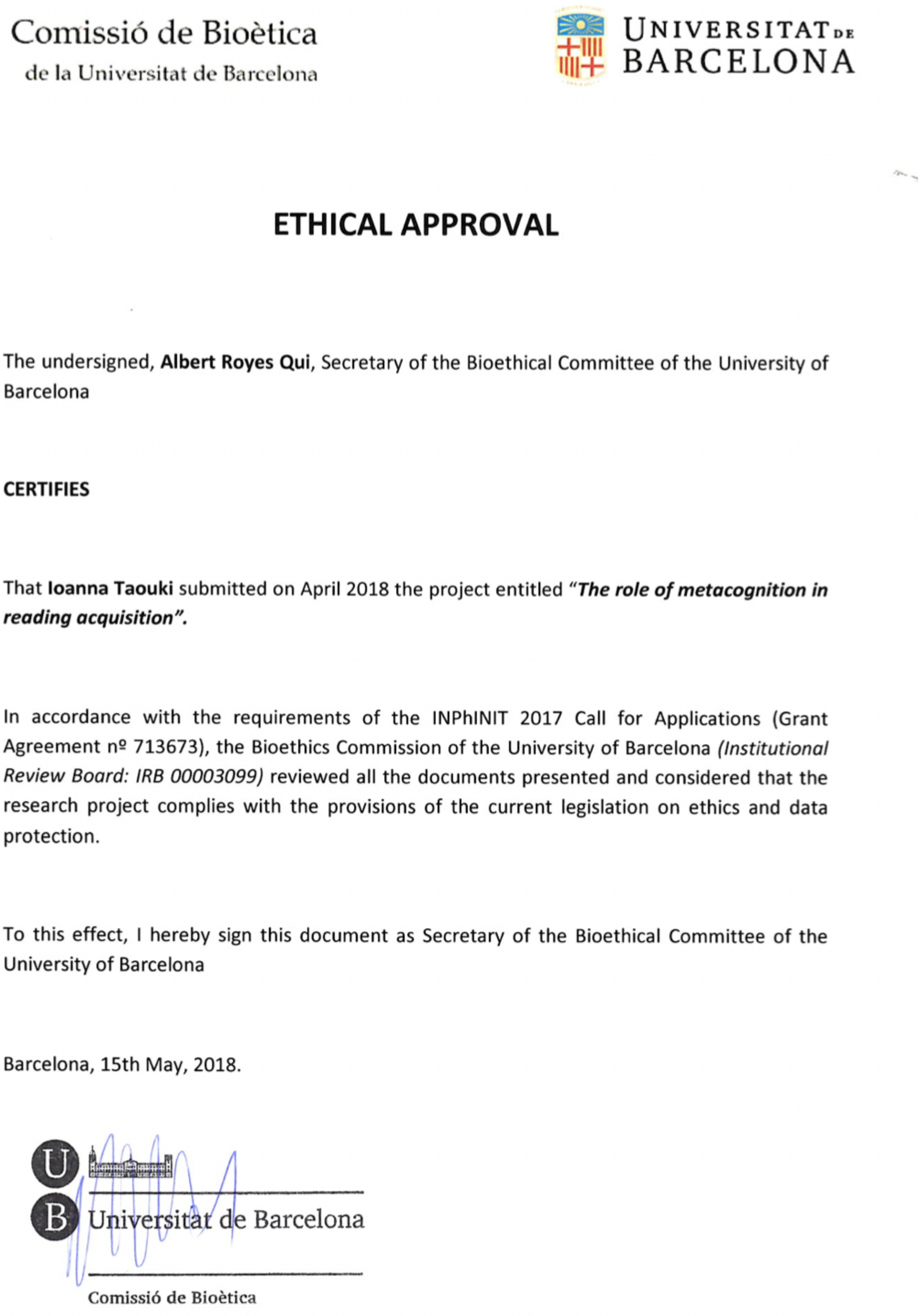
What happens if I start helping and then I don't want to take part anymore?

Don't worry if at any time you decide to stop participating! You are allowed to do that and nobody will be angry about it!

Your signature or Name ______________________________

10. Appendix B

10.1. B1. Ethical assessment of the present study



Comissió de Bioètica
de la Universitat de Barcelona

UNIVERSITAT DE BARCELONA

ETHICAL APPROVAL

The undersigned, **Albert Royes Qui**, Secretary of the Bioethical Committee of the University of Barcelona

CERTIFIES

That **Ioanna Taouki** submitted on April 2018 the project entitled ***“The role of metacognition in reading acquisition”.***

In accordance with the requirements of the INPhINIT 2017 Call for Applications (Grant Agreement nº 713673), the Bioethics Commission of the University of Barcelona *(Institutional Review Board: IRB 00003099)* reviewed all the documents presented and considered that the research project complies with the provisions of the current legislation on ethics and data protection.

To this effect, I hereby sign this document as Secretary of the Bioethical Committee of the University of Barcelona

Barcelona, 15th May, 2018.

UB Universitat de Barcelona
Comissió de Bioètica

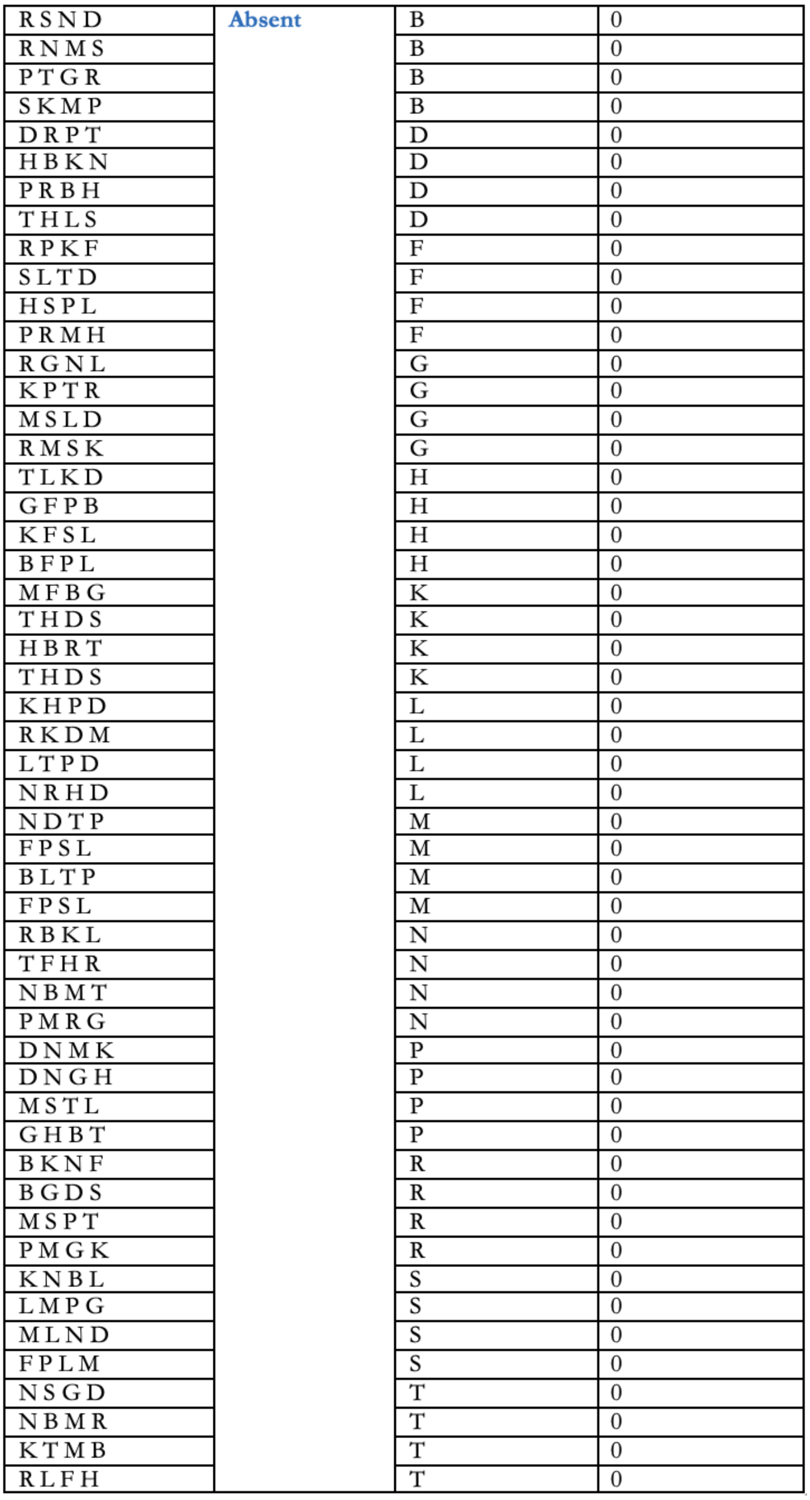
11. Appendix C

11.1. C1. Visual Attention Span Task stimuli

Table C1 Stimuli used in the VA span task.

Letter sequence	Trial Type	Target	Position
B G D R	Present	B	1
D T P S		D	1
F N M L		F	1
G T L R		G	1
H M S P		H	1
K B G H		K	1
L M G P		L	1
M B P T		M	1
N T D S		N	1
P N K T		P	1
R F P N		R	1
S T B L		S	1
T S N L		T	1
G B T D		B	2
K D P N		D	2
D F R K		F	2
M G N B		G	2
F H M R		H	2
N K H M		K	2
F L D B		L	2
G M H B		M	2
H N T K		N	2
B P F R		P	2
M R L G		R	2
M S L H		S	2
L T D N		T	2
M G B H		B	3
G B D H		D	3
H T G K		F	3
R K F T		G	3
T N H P		H	3
S N K P		K	3
K D L P		L	3
F D N H		M	3
N R M K		N	3
H S P F		P	3
N K R F		R	3
G N S P		S	3
B G T F		T	3
H L R B		B	4
S F P D		D	4
L R N G		F	4
N G B F		G	4
M L B H		H	4
S G F K		K	4
F B R L		L	4
S P L N		M	4
F T H D		N	4
K H F P		P	4
G S B R		R	4
T P L S		S	4
B G K T		T	4

Table C1 - (continuation)



RSND	**Absent**	B	0
RNMS		B	0
PTGR		B	0
SKMP		B	0
DRPT		D	0
HBKN		D	0
PRBH		D	0
THLS		D	0
RPKF		F	0
SLTD		F	0
HSPL		F	0
PRMH		F	0
RGNL		G	0
KPTR		G	0
MSLD		G	0
RMSK		G	0
TLKD		H	0
GFPB		H	0
KFSL		H	0
BFPL		H	0
MFBG		K	0
THDS		K	0
HBRT		K	0
THDS		K	0
KHPD		L	0
RKDM		L	0
LTPD		L	0
NRHD		L	0
NDTP		M	0
FPSL		M	0
BLTP		M	0
FPSL		M	0
RBKL		N	0
TFHR		N	0
NBMT		N	0
PMRG		N	0
DNMK		P	0
DNGH		P	0
MSTL		P	0
GHBT		P	0
BKNF		R	0
BGDS		R	0
MSPT		R	0
PMGK		R	0
KNBL		S	0
LMPG		S	0
MLND		S	0
FPLM		S	0
NSGD		T	0
NBMR		T	0
KTMB		T	0
RLFH		T	0

11.2. C2. Orthographic Lexical Decision Task stimuli

Table C2 Stimuli used in the lexical decision task.

Item	Lexicality	Time point of the study	Word frequency (log)	Item structure
CASA	**Word**	**Grade 1**	2.8	CVCV
AGUA			2.47	VCVV
AMOR			2.43	VCVC
HIJO			2.35	CVCV
CARA			2.35	CVCV
AIRE			2.32	VVCV
IDEA			2.29	VCVV
MESA			2.24	CVCV
BOCA			2.19	CVCV
BASE			1.96	CVCV
CABO			1.96	CVCV
DUDA			1.93	CVCV
SALA			1.93	CVCV
COPA			1.69	CVCV
ROSA			1.68	CVCV
COLA			1.58	CVCV
FOTO			1.57	CVCV
VOTO			1.51	CVCV
LEÓN			1.49	CVVC
BESO			1.48	CVCV
TELE			1.39	CVCV
SACO			1.36	CVCV
RAZA			1.33	CVCV
RUTA			1.31	CVCV
GOTA			1.3	CVCV
FILA			1.28	CVCV
SOFÁ			1.27	CVCV
TAZA			1.25	CVCV
NUBE			1.25	CVCV
RAMA			1.24	CVCV
MOZO			1.2	CVCV
PUÑO			1.18	CVCV
LANA			1.18	CVCV
BOLA			1.18	CVCV
GOMA			1.17	CVCV
GOZO			1.15	CVCV
ROBO			1.14	CVCV
POLO			1.14	CVCV
VELO			1.11	CVCV
VACA			1.08	CVCV
MADO	**Pseudoword**	**Grade 1**	2.59	CVCV
COGA			2.53	CVCV
HONA			2.34	CVCV
NIPO			2.29	CVCV
OBLA			2.21	VCCV
EMAD			2.19	VCVC
TEÑA			2.13	CVCV
CIME			2.1	CVCV
ZOSA			2.07	CVCV
JEÑE			2.03	CVCV
PEJO			1.95	CVCV

Table C2 - (continuation).

AZTO	**Pseudoword**	**Grade 1**	1.9	VCCV
PIGO			1.85	CVCV
MIPO			1.63	CVCV
FADA			1.62	CVCV
GACO			1.59	CVCV
VAPO			1.58	CVCV
DASO			1.47	CVCV
NAPE			1.46	CVCV
SEBA			1.4	CVCV
BOMA			1.38	CVCV
TAÑI			1.37	CVCV
CEMO			1.31	CVCV
RIDA			1.3	CVCV
AUDO			1.27	VVCV
LAFO			1.26	CVCV
ALRA			1.26	VCCV
CEMA			1.23	CVCV
SOZA			1.2	CVCV
TUJO			1.2	CVCV
NIRA			1.19	CVCV
REGO			1.19	CVCV
PAFA			1.18	CVCV
RAUZ			1.16	CVVC
PIMA			1.06	CVCV
BOPE			1.06	CVCV
NISO			1.04	CVCV
MOSO			1.03	CVCV
BOBE			1	CVCV
COSA	**Word**	**Grade 3**	2.53	CVCV
OBRA			2.21	VCCV
CINE			2.1	CVCV
FRÍO			2	CCVV
CAFÉ			1.89	CVCV
ROJO			1.81	CVCV
PAPÁ			1.67	CVCV
FAMA			1.62	CVCV
CITA			1.59	CVCV
DATO			1.47	CVCV
HOJA			1.43	CVCV
TAXI			1.37	CVCV
CERO			1.31	CVCV
POZO			1.25	CVCV
CERA			1.23	CVCV
TUBO			1.2	CVCV
PIPA			1.19	CVCV
RAÍZ			1.16	CVCV
NIDO			1.04	CVCV
RIMA			0.44	CVCV
HORA			2.34	CVCV
EDAD			2.19	VCVC
ZONA			2.07	CVCV
PESO			1.95	CVCV
RATO			1.87	CVCV
MASA			1.77	CVCV

Table C2 - (continuation).

HUMO	**Word**	**Grade 3**	1.66	CVCV
CURA			1.6	CVCV
GATO			1.59	CVCV
NAVE			1.46	CVCV
SEDA			1.4	CVCV
MODA			1.36	CVCV
ONDA			1.3	VCCV
ALBA			1.26	VCCV
LOBO			1.21	CVCV
VELA			1.2	CVCV
RETO			1.19	CVCV
BOTE			1.06	CVCV
MOTO			1.03	CVCV
RAMO			0.99	CVCV
PAÉS	**Pseudoword**	**Grade 3**	2.49	CVVC
CAWA			2.35	CVCV
IPEA			2.29	VCVV
LOGA			2.19	CVCV
PEBO			2.01	CVCV
DUCA			1.93	CVCV
ELOR			1.86	VCVC
MOBA			1.69	CVCV
RIPA			1.6	CVCV
MOMA			1.58	CVCV
VOMO			1.51	CVCV
MUMO			1.44	CVCV
MAVA			1.37	CVCV
RURA			1.31	CVCV
FISA			1.28	CVCV
NUPE			1.25	CVCV
MOPO			1.2	CVCV
PUBO			1.18	CVCV
TOLO			1.15	CVCV
VENO			1.11	CVCV
AGUM			2.47	VCVC
HIVO			2.35	CVCV
TADO			2.27	CVCV
SIEL			2.11	CVVC
BADE			1.96	CVCV
SABA			1.93	CVCV
DESO			1.72	CVCV
ROVA			1.68	CVCV
ISGA			1.59	VCCV
JOTO			1.57	CVCV
REÓN			1.49	CVVC
TEBE			1.39	CVCV
SAÑO			1.36	CVCV
GOCA			1.3	CVCV
PONO			1.28	CVCV
TAVA			1.25	CVCV
FANA			1.18	CVCV
GOPA			1.17	CVCV
PEMO			1.14	CVCV

11.3. C3. Emotion Recognition Task stimuli

Table C3 Stimuli used in the emotion recognition task.

Table C3 - (continuation).

11.4. C4. Standardized Reading Tests stimuli (PROLEC)

Table C4-1 Word stimuli used in the standardized reading tests (example trials in gray).

List of Words

casa	barco	prado	
globo	peine	pueblo	ciervo
ermita	fuego	gigante	cuerpo
girasol	especie	treinta	granizo
ombligo	trono	blanco	alfombra
pulga	trompeta	prensa	viento
huelga	muerto	lienzo	cristal
estrella	mueble	princesa	astuto
bosque	sombrero	tierra	cloro
peldaño	gente	triunfal	plato
tintero	liebre	pregunta	tractor

Table C4-2 Pseudoword stimuli used in the standardized reading tests (example trials in gray).

List of Pseudowords

reca	tispe	blopa

gloro	peima	pueña	ciergo
erpisa	fueme	giranco	cuerla
gicamol	escodia	treindo	graliza
onclaso	trollo	blansa	almiento
pulda	trondeja	prencol	vienca
huelte	muerbo	lienca	crispol
escrilla	muepla	prinsota	ascuso
bospe	sodiro	tiepre	clofo
pelcafo	genso	triundol	plafo
tincoro	liegra	prejonta	tractan

<u>Note:</u> The cover image of this thesis is a composite of drawings made by members of my family and close friends. At the end of my thesis, I asked them to draw something related to what they had understood my thesis was about. To me, this serves as proof that in science, as in life, everyone has a different perspective that is equally valuable, but through teamwork, we can better understand the world surrounding us.

BASQUE CENTER
ON COGNITION, BRAIN
AND LANGUAGE