eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Reading acquisition: from digital screening to neurocognitive bases in a transparent orthography

Camila Zugarramurdi

Under the supervision of
Dr. Juan Valle Lisboa & Dr. Manuel Carreiras

This dissertation is submitted for the degree of

Doctor of Philosophy

June 2021

Abstract

The first problem: Current practices in detection of reading difficulties

Reading difficulties carry cascading consequences for socioemotional and professional development (Arnold et al., 2005). Current practices in detection of reading difficulties typically entail waiting until 3rd grade elementary school to arrive at a proper diagnosis and offer remedial interventions (Ozernov-Palchik & Gaab, 2016). However conservative this strategy is, which is a safeguard for avoiding overdiagnosis and false positives, its practical consequences are too damaging to sustain it. For decades now there has been some knowledge to predict, with more or less sensitivity, future reading trajectories from assessing a few critical skills that develop prior to formal reading instruction (see for example, Lyytinen et al., 2015). Nevertheless, this knowledge has not led to adequate prevention strategies. In addition, it is currently well known that remedial practices are most effective the earlier they begin (Wanzek & Vaughn, 2007). Thus, the present wait-for-failure approach has dire consequences. A comprehensive screener for identifying children at risk of presenting reading difficulties when they are in kindergarten is presented in Study One. This screener is classroom-based, digital, self-administered, and brief, making it feasible and cost-effective. Results show that only four variables (letter knowledge, phonological awareness, non-verbal short-term memory, and socioeconomic status) are sufficient to attain high levels of classification accuracy. Therefore, screening can and should be done universally in kindergarten to prevent some of the reading failure trajectories.

The second problem: Current issues in reading universals

The universality of the cognitive substrates of reading performance is currently under debate. The majority of studies have been carried out with English speakers (Share, 2008), an alphabetic script with an inconsistent orthography (i.e., the mapping between graphemes and phonemes is not consistent). In Study Two we investigate to what extent can findings from inconsistent orthographies be translated into a consistent orthography such as Spanish. Not only because of the need to produce local knowledge to address local issues, but also to enrich the academic discussion through novel, infrequent, evidence. The database obtained from the comprehensive screener presented in Study One enables thorough control of confound variables in a longitudinal model of reading outcomes. Results show that the role of preliteracy skills differs in Spanish when compared to other less consistent orthographies. In particular, phonological awareness does not seem to contribute to reading acquisition above and beyond other preliteracy skills. In turn, letter knowledge takes a more central role in the prediction of early reading skills. We propose that a delayed developmental trajectory for phonological

awareness —influenced by home literacy and educational practices as well as the intrinsic characteristics of the orthographic system—, a strong role for verbal short-term memory, and a tighter association between letter knowledge and phonological awareness can accommodate this findings.

The third problem: cognitive basis of phonological processing

Phonological processing has been at the core of the mechanistic understanding of reading difficulties. While a handful of theories explain reading difficulties through additional mechanisms (visuo-spatial, attentional, motor, and in statistical learning or anchoring), phonology's role is undisputed. It stands to question then what the sensory-cognitive basis of phonological processing are. A strong candidate since early days of phonological processing studies is rhythm. Rhythmic sensitivity underlies processing of stress in speech, which in turn plays a central role in speech parsing and segmentation, which underly the formation of phonological representations. In Study Three, we investigate the connection between rhythmic skills, phonological processing, and future reading outcomes. Results show that rhythmic sensitivity predicts future reading acquisition through phonological processing but also above and beyond it. We discuss the implications of the role of rhythmic processing for reading acquisition.

The fourth problem: neural underpinnings of reading performance

If rhythm is central to speech segmentation, phonological processing and thus reading, as shown in Study Three, then auditory processing of rhythm at the neural level should underly phonological processing and relate to reading. While many studies have addressed this issue in adults and older children, barely a handful have studied it in prereaders. Thus, in Study Four we examine the neural processing of auditory rhythmic stimuli, and its relation to future reading acquisition. Results show that neural responses in prereading children synchronize to auditory rhythms, and that this synchronization is related to future reading outcomes. This finding provides novel evidence on the role that cortical oscillations play in auditory processing and reading acquisition.

When we were hunter-gatherers, language became essential for planning the day's activity, teaching the children, cementing the friendships, alerting the others to danger, and sitting around the fire after dinner watching the stars come out and telling stories. Eventually, we invented phonetic writing so we could put our sounds down on paper and, by glancing at a page, hear someone speaking in our head —an invention that became so widespread in the last few thousand years that we hardly ever stop to consider how astonishing it is.

Carl Sagan, 1998, p. 42.
Billions and Billions: Thoughts on Life
and Death at the Brink of the Millennium

Contents

List of Figures

List of Tables

Credits

El CIBPsi (Uy)
(in clustering order)

Juan Valle Lisboa,
on the verge of our 10
year anniversary and
still liking each other
Alvaro Cabana,
mixed models guru
and much more
Emilia Fló,
a great future awaits us
Lucía Fernández,
what a journey,
made even better by
your riding along
Soledad Assis,
a learning opportunitiy
in every encounter
The Lexiland team &
Mercedes Techera for
their comittment
Alejandro Maiche,
the chess player
Alejandra Carboni,
for roberto quenedi
Hernán Delgado,
better late than never
Verónica Nin
María Noel Macedo
Everyone

The BCBL (Sp)
(in clustering order)

Manuel Carreiras,
for the opportunity
and his doors always open
Marie Lallier,
the reading expert
Landis group
Alexia Antzaka, *the* person to
discuss science, and more
Candice Frances, editor and
proofreader #1, and compatriota
Irene Monsalve
Usman Ayub Sheikh
Borja Blanco
Yuriem Fernández
Brendan Costello
Alejandro Perez
Eider Juaristi & Ana Fernández,
constant support
Old predoc gang,
for drinks and pintxos
Young predoc gang

BCBL family
(in order of appearance)

Sophie Schloeffel,
delicious company
Joyse Medeiros, pure wit
Patricia Alves Dias,
so much laugh
Noemi Fariña, essential hugger
Jaione Arnaez, true basque friend

The old family
(in order of appearance)

Mabel García,
endless support
Mario Zugarramurdi,
constant pillar
Javier Zugarramurdi,
creative outcast
Maite Zugarramurdi,
ultimate carer
Ana Zugarramurdi.
shoulder and buddy
Martín Zugarramurdi,
best twin and better manny

The new family
(equal contribution)

Agustín Dieste,
relentless companion, soul
Feliciano Dieste Zugarramurdi,
magic love

Funding
Fundación Carolina (Sp)
Agencia Nacional de Investigación e Innovación (Uy)

Chapter 1

Background

Reading. It is a means for pleasure, work, learning, sharing, and remembrance. Reading is aimed at conveying meaning. The ultimate goal of the act of reading is comprehension. But comprehension does not come easily. Unlike language, reading is an acquired skill that takes instruction and practice for mastery. Our organisms evolved in an environment where writing did not exist. Thus, neural circuits for reading are not prewired, they need to be built *de novo* in every individual facing the challenge of learning to read. How is this feat accomplished? It seems quite clear today that reading is built upon basic oral language skills that start developing very early during child development: vocabulary, grammar, speech processing, among others. In the following sections we go through some of the things we know about reading acquisition, and some of the things we need to understand better.

1. Reading acquisition

Various models of reading acquisition converge on four distinct, but overlapping, phases (see Ehri, 2008 for an overview). In the first phase, children "read words" by recognizing salient visual cues but not by alphabetic knowledge. This is the case with frequently seen signs, brands, or their own name. The second phase is *decoding,* which entails developing the mechanics of reading. Decoding consists of sounding out each letter in a written word, and blending them together in order to attain the word's pronunciation and corresponding identity (Scarborough & Brady, 2002). By doing so, children can "hear themselves" pronouncing a word, which takes them from the new visual code to the phonological code with which they are already familiar (which subsequently grants access to the word's semantic information). With enough practice, this processed is *automatized*: instead of sounding out each letter, children can now visually recognize familiar words quickly and access their meaning. When high accuracy levels for familiar words have been achieved, *fluency* and *comprehension* start emerging (Kuhn & Stahl, 2003). Fluency is characterized by increasing levels of reading speed at the word and text level, while maintaining adequate levels of accuracy. The accompanying liberation of cognitive resources, no longer directed towards decoding, paves the way for increasing comprehension (Abusamra et al., 2008). Comprehension is of course the ultimate goal of reading. Successful reading comprehension depends on having achieved automatic visual word recognition (and thus fluency), and also from vocabulary and higher level processes such as inference construction and metacognitive skills (Castles et al., 2018).

In expert readers, the end state of visual word recognition can be observed. According to computational, cognitive and neuroimaging evidence, word meaning can be accessed from two different sources, depending on word frequency and regularity (Coltheart et al., 2001; Harm & Seidenberg, 2004; Jobard et al., 2003; Zoccolotti et al., 2005). In a direct or lexical pathway, word meaning is accessed directly from its written representation. This is possible since repeated encounters with written words during early reading support the development of an orthographic lexicon, which provides a direct mapping from the word's printed form to its meaning (McCandliss et al., 2003). The alternative indirect or sublexical pathway, used mainly for infrequent words, still uses the slow and laborious process of decoding, turning each letter into its corresponding sound (i.e., grapheme to phoneme conversion), blending them together and, via its phonological form, access word's meaning. This indirect, sublexical pathway is not useful for words with irregular spelling, where its phonological form cannot be obtained by applying grapheme to phoneme correspondences. In such cases, only a lexical route can be counted on. This is particularly relevant for English, where a large number of words are irregular. In Spanish, on the contrary, almost all words' phonological forms can be accessed through applying grapheme to phoneme conversion rules (this topic will be further discussed in Section 1.2).

The indirect, subexical route is predominant in the decoding phase during reading acquisition (Castles et al., 2018; Turkeltaub et al., 2003; Zoccolotti et al., 2005). It is via repeated and systematic practice with word decoding that early readers can move on into the next phases. Thus, while decoding is not sufficient for attaining reading comprehension, it is a necessary condition. Therefore, decoding serves a fundamental role in reading acquisition. In turn, what skills are necessary for successful decoding? The past four decades of research in reading acquisition have pointed to a set of foundational skills that develop before the onset of reading acquisition (henceforth called *preliteracy skills*). In the following section we describe them in more detail.

1.1. Preliteracy skills

Across studies, three preliteracy skills have been identified: knowledge of letter sounds (letter knowledge; LK); rapid and efficient access to lexical representations (rapid automatized naming; RAN); and an ability to consciously manipulate the constituent units of oral language, generally referred to as phonological awareness (Boets et al., 2006; Lyytinen et al., 2006b; Muter et al., 2004; Schatschneider et al., 2004). A fourth skill that taps into phonological processes—and also predicts early reading acquisition—but has received far less attention is verbal short-term memory.

1.1.1. Phonological awareness

Among these preliteracy skills, phonological awareness has been identified as the most central (Vellutino et al., 2004). *Phonological awareness* can be defined as the conscious access and deliberate manipulation of phonological representations. *Phonological representations* are "mentally represented information about the phonological characteristics of a particular word"

(Scarborough & Brady, 2002, p. 306). These may refer to any of the constituent units of words: syllables, onset-rime, or phonemes (see section 1.2 for a definition of phonemes). When phonological awareness involves manipulation of phonemes, it is often referred to as *phonemic awareness,* although sometimes these two terms are used interchangeably. Phonological awareness is commonly measured through tasks involving segmenting words, blending phonemes or syllables to form a word, identifying phonemes within a word, or indicating whether two words rhyme. Although it is generally considered a unidimensional construct, different tasks tap into subtle different processing mechanisms. In general, tasks involving synthesis (such as blending) are easier than tasks involving analysis (such as identifying sounds withing a word) (Torgesen et al., 1994). With respect to its development, studies have shown that phonological awareness proceeds from larger to smaller grain sizes, i.e., from syllables to phonemes (Anthony et al., 2003). An additional important feature of phonological awareness *tasks* is the production of a verbal response. While the phonological awareness construct does not include a verbal response at its core, almost all tasks used to measure phonological awareness do. It has been shown that the production of a verbal response contributes to the phonological awareness-reading association, above and beyond the contribution of the actual phonological awareness manipulation (Cunningham et al., 2015).

A large body of evidence on the role of phonological awareness in reading acquisition comes from studies involving children with reading difficulties, i.e., *dyslexia*. When compared with typical readers, dyslexic children have shown poorer performance in phonological awareness tasks (see Melby-Lervåg et al., 2012 for a meta-analysis). Moreover, intervention studies have shown that training phonological awareness skills improves reading acquisition (Hulme et al., 2012; Vellutino & Scanlon, 2013). This body of evidence has led to the phonological deficit hypothesis in dyslexia, which posits that phonological deficits are the cause of reading difficulties in most dyslexic children (but see Ramus & Szenkovits, 2008). Why would phonological awareness be so important for reading acquisition? Because decoding consists of mapping letters to their corresponding phonemes, and for such a mapping to take place, emergent readers need to access their phonological representations, i.e., possess phonological awareness. Accordingly, phonological awareness has been shown to be a main predictor of decoding skills, while playing a less important role in attaining fluency or comprehension (Muter et al., 2004).

1.1.2. Letter knowledge

The second preliteracy skill identified as foundational to reading acquisition is possessing knowledge of letters names and/or sounds. Children frequently learn letter names or sounds before entering kindergarten, at their home environment from books, songs, magnets, or wooden cubes (Seidenberg, 2017). By being exposed to print material, children start grasping the alphabetic principle. The alphabetic principle is the notion that print represents sounds. Children need to understand that this is the basis of the writing system. Directing children's attention towards how letters map to phonemes aids in this understanding (Foulin, 2005).

A recurrent issue in the literature is whether letter names or letter sounds should be taught, and how these two relate to each other. While letter sounds are directly relevant for mapping each letter into its corresponding phoneme during decoding, letter names are not. This has led some researchers to propose that letter names should not be taught (see a discussion of this topic in Foulin, 2005). However, available evidence suggests that learning letter names aids in learning letter sounds, especially when letter names begin with the sound of the letter, for example in <b>, but not in <m> (Cardoso-Martins et al., 2011; Treiman et al., 1996). One explanation for this association is that letter names work as labels which aid in categorization (Lupyan et al., 2007). The rationale is the following. Letters are represented by visual patterns that vary greatly. For example, the letter <a> can be encountered in handwriting (which varies from person to person), in print (which varies from font to font), in lowercase or uppercase, among others. Thus, children need to learn that all these different visual forms a letter can take compose a single category, i.e., the letter <a>. In principle, this is categorization problem similar to understanding that all chairs, independently of their particular features, correspond to a single conceptual category, a *chair*.

Letter knowledge and phonological awareness are tightly related. In order to learn letter sounds, the child needs to have access to some form of abstract representation of the sounds in the language, i.e., phonological awareness. At the same time, knowledge of letters names and sounds aids in the development of phonological awareness (Cardoso-Martins et al., 2011; Foulin, 2005)

1.1.3. RAN

Rapid automatized naming is a task rather than a skill. The task involves speeded naming of a grid of items representing either colours, objects, letters, or numbers. It was first developed as a means of assessing the association between naming and reading observed in patients with alexia, an acquired deficit with reading. Since then, it has been thoroughly used in reading research, showing that dyslexic children show poor performance in the task (see Araújo et al., 2015 for a meta-analysis). What are the underlying cognitive skills that the RAN task measures? At the item level, it assesses automatization in retrieving lexical items from memory (i.e., lexical access or retrieval). In this sense, RAN reflects phonological processing skills. Additionally, at the supra-item level, it indexes the visuo-attentional skills necessary to process series of stimuli, since items are arranged in series. Indeed, RAN has been shown to predict performance in both decoding and fluency, reflecting both the item and supra-item skills that the task involves (Norton & Wolf, 2012).

The phonological aspect of the RAN tasks has questioned whether it should be considered as part of the core phonological deficit in dyslexia, or whether it should be considered as an independent source of reading difficulties (e.g., Ramus & Szenkovits, 2008; Torgesen et al., 1994). Based on several studies showing dissociations between RAN and phonological skills, the double-deficit hypothesis posited that these two represent independent sources of reading difficulties, and that when both coincide in the same child, reading difficulties are more severe (Bowers & Wolf, 1999).

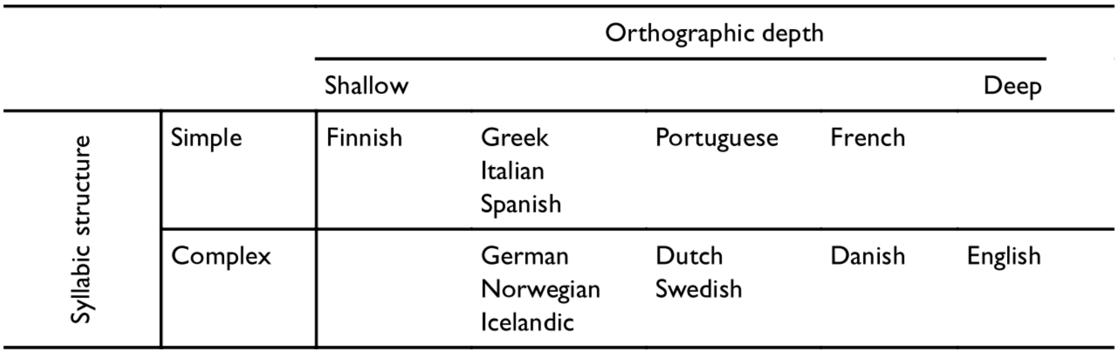
1.1.4. Verbal short-term memory

Verbal short-term memory (also known as phonological short-term memory or auditory short-term memory) refers to the ability to temporarily store phonological information in memory. It is typically assessed through repeating a series spoken items, usually digits or nonwords, of increasing length. The reason for preferring digits or nonwords instead of words lies on trying to exclude any aid coming from semantic processing of the stimuli, which could differ among participants and thus act as a confound. The underlying processes in verbal short-term memory can be distinguished in at least two dimensions. On one hand, with respect to the type of memory, two types —often lumped together— can be distinguished: memory for items and memory for order. On the other hand, with respect to the processing involved, we can distinguish the encoding stage from the retrieval stage. Finally, it is important to distinguish verbal short-term memory from working memory. Unlike working memory, verbal short-term memory does not involve manipulating the stored information, just recalling it in the same order as presented.

Verbal short-term memory has received a mixed treatment in the reading literature, sometimes treated as a foundational skill of interest in reading acquisition, together with other preliteracy skills (Moll et al., 2014; Ramus & Szenkovits, 2008; Torgesen et al., 1994), and sometimes as a covariate (Caravolas et al., 2012; Furnes & Samuelsson, 2010; Puolakanaho et al., 2007; Vaessen et al., 2010). The latter responds to the fact that almost all phonological awareness tasks include a substantial verbal short-term memory component. Therefore, in order to isolate the phonological awareness component from the memory component, it is important to control for memory skill. Verbal short-term memory and phonological awareness together can be considered as two separate components of *phonological processing.* Arguably, when assessing the contribution of phonological processing to reading acquisition, verbal short-term memory has received much less attention than phonological awareness.

1.2. Orthographic consistency

As explained before, one central step in decoding is sounding out each letter in a written word and blending them together in order to attain its pronunciation and identity. This process depends on the knowledge of *grapheme*-to-*phoneme* mappings. Phonemes are abstract representations of the minimal sound units of language, and graphemes are its written counterpart. Graphemes can consist of single letters, or of a combination of letters. For example, the grapheme <sh> in English, or <ll> in Spanish, uses two letters to represent a single phoneme. In general, though, one grapheme corresponds to one letter, and these two terms are used interchangeably.



		Orthographic depth				
		Shallow				Deep
Syllabic structure	Simple	Finnish	Greek Italian Spanish	Portuguese	French	
	Complex		German Norwegian Icelandic	Dutch Swedish	Danish	English

Figure 1.1. Orthographic consistency across languages. From Seymour, Aro, & Erskine, 2003.

Orthographies differ in the consistency of their grapheme-to-phoneme mappings, usually classified in a continuum from transparent or shallow (high consistency) to opaque or deep (low consistency) (Figure 1.1). English is classified at the opaque end of this continuum. In English, grapheme-to-phoneme mappings are very inconsistent. For example, the letter <a> in the words *cat* and *table* correspond to very distinct phonemes (/'kæt/ and /'teɪbəɫ/, respectively), as the letter <u> in the words *put* and *but* (/'pʊt/ and /'bət/, respectively). Spanish is classified on the other end of the continuum. In Spanish, grapheme-to-phoneme mappings are very consistent, with the few exception of <g> and <c>. For example, the letter <g>, can sound weak or strong depending on the context as in *mago* (magician) and *mágico* (magical) (/'ma.ɣo/ and /'ma.xi.ko/, respectively). Indeed, Spanish and English share the same number of letters in the alphabet but use different numbers of phonemes in language. Whereas in Spanish 26 letters are used to represent 25 phonemes —which yields an almost one-to-one mapping—, in English the same 26 letters are used to represent approximately 36 phonemes —which yields a one-to-many mapping. Thus, in Spanish, knowing letters sounds is almost equivalent to knowing the phonemes in the language. In contrast, in English, knowledge of canonical letter sounds (those regularly taught at home or in school) represents only a subset of the phonemes in the language. Furthermore, the inconsistency of the English orthographic system deems grapheme-to-phoneme correspondences an insufficient strategy for word decoding, since letter sounds cannot be inferred from context (for example in the two homophones in *wind*, one rhyming with *pinned* and another rhyming with *find* (Van Orden & Kloos, 2005)). In such cases, children need to resort to larger grain sizes, such as morphology, in order to achieve correct pronunciation and visual word recognition.

In sum, children learning to read in an inconsistent orthography need to learn more grapheme-to-phoneme correspondences, and to resort to additional decoding strategies, while children learning to read in a consistent orthography are almost ready to decode once they have mastered letter sound knowledge. Indeed, it has been repeatedly shown that orthographic consistency modulates the rate of reading acquisition, with higher accuracy levels in decoding being achieved faster in more transparent orthographies (Mann & Wimmer, 2002; Moll et al., 2014; Seymour et al., 2003).

1.3. Summary

Reading acquisition proceeds from decoding, to fluency, to comprehension. The success of this pathway relies —among other things— on developing fast automatic decoding skills, which in turn depend on possessing strong preliteracy skills, namely phonological awareness, letter knowledge and rapid automatized naming. Orthographic consistency likely modulates this process, by changing how useful different reading pathways are, and, in turn, the relative contribution that preliteracy skills have across orthographies. As a way of understanding individual differences in preliteracy skills, we next turn onto phonological development.

2. Phonological development

Phonology is the subsystem of language that refers to how speech represents language, and the knowledge that a speaker has of the sound properties of speech (Ingram, 2007; Scarborough & Brady, 2002). Among aspects of phonology, *phonological representations* are key to reading acquisition. Phonological representations are the abstract representation of speech sounds. These entail the word, syllable, and phoneme level, as well as information on stress and articulation. The access to these phonological representations is what we have termed phonological awareness. In order to attain phonological representations at different grain sizes (e.g., lexical, syllabic, and phonemic), infants need to segment the speech input, an otherwise continuous signal. How is speech segmentation achieved? And how are the different grain sizes integrated? Rhythms in speech and in brain activity seem to provide an answer.

2.1. Rhythms in speech

At the acoustic level, speech is conveyed through continuous modulations in sound pressure (Figure 1.2). Speech is composed of different *rhythms,* corresponding to amplitude modulations at different temporal scales, corresponding broadly to prosodic, syllabic, and phonemic information. These rhythms are hierarchically nested. Phonemic information falls within a temporal window corresponding to the syllable level, and syllabic information falls within a temporal window corresponding to the phrasal or prosodic level.

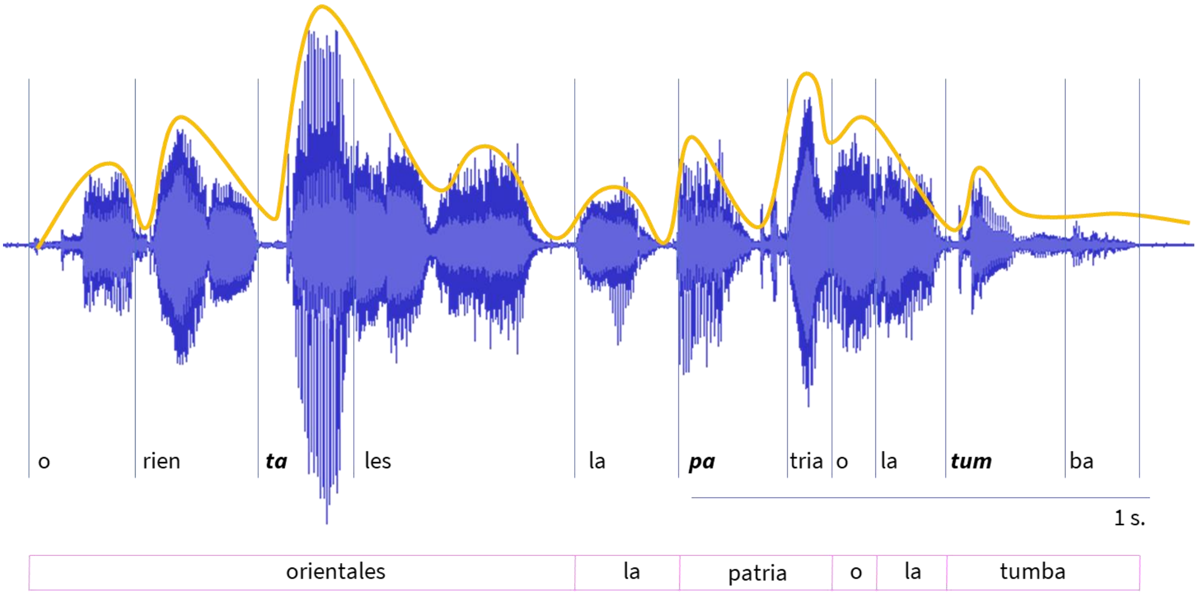


Figure 1.2 Acoustic waveform for the phrase "orientales la patria o la tumba" [orientales, our nation or the grave]. The speech envelope (orange line), the sum of the energy profiles at each modulation frequency, reflects syllabic rate, at approximately the theta rate (4 - 8 Hz).

During speech processing, the continuous acoustic signal needs to be segmented and mapped into meaningful linguistic units. However, speech segmentation is a non-trivial problem since, as can be seen from Figure 1.2, no clear acoustic boundaries mark the end of a word or phoneme and the beginning of the next one. In contrast, clear acoustic markers signal syllable boundaries in most syllables These acoustic markers conform *speech edges*, transient sharp events in the modulation of the waveform. Given its saliency, syllabic rhythm dominates the hierarchy, as reflected in the speech envelope.

Early foundational work by Anne Cutler and Jacques Mehler proposed that speech segmentation is based on the analysis of the rhythmic structure of speech (Cutler et al., 1986; Cutler & Mehler, 1993; Mehler et al., 1981). The general idea is that a listeners' attention is biased towards stress in speech, which aids in the segmentation of the continuous speech stream. In English, most words have a stress pattern in which the first syllable is stressed (strong) followed by an unstressed (weak) syllable. Thus, listeners can use lexical stress as a cue for word boundaries. Accordingly, infant studies have shown that sensitivity to stress develops over time, and that at 7 months old, infants use strong/weak patterns in words as cues to word boundaries (Jusczyk et al., 1999). Under this framework, stress sensitivity is therefore an index of speech segmentation abilities and underlies the development of phonological representations.

However, this leaves unattended the problem of how phonological representations at different grain sizes arise from speech segmentation, and specially how they are integrated. The hierarchical structure in speech and in brain rhythms have provided some answers.

At the neural level, accounts of speech processing through cortical oscillatory activity have shed light on how speech segmentation at different grain sizes is achieved and integrated. Cortical oscillations are coordinated fluctuations in the electrical activity of ensembles of neurons. In the auditory cortex, oscillations are mainly observed at three frequency bands: delta (0.5 – 3 Hz), theta (4 – 8 Hz), and gamma (above 30 Hz). Similarly, in speech, information at different grain sizes occurs in equivalent time scales: prosodic at delta, syllabic at theta and phonemic at gamma rate. It is possible then that these matching rhythms between brain and speech are exploited by the brain to track and segment the speech signal. Indeed, experimental studies have found speech-brain synchronization at these frequencies (Meyer, 2018; Peelle & Davis, 2012).

Currently, the underlying synchronization mechanisms are being thoroughly studied (Gourévitch et al., 2020; Haegens & Zion Golumbic, 2018). In a seminal study (Gross et al., 2013), cortical delta and theta oscillations were shown to modulate their *phase* to speech edges (phase-reset). At theta rate, synchronization was observed mainly at temporal auditory areas, whereas at delta rate synchronization spread more anteriorly towards right frontal areas. This was interpreted as theta and delta subserving qualitatively different functions, with delta involved in processing prosodic features of the speech signal, and theta involved in syllabic processing. In turn, cortical gamma oscillations were shown to modulate their *amplitude* to speech edges. As explained above, speech edges, observed in the speech envelope, occur approximately in the theta range, corresponding to the change rate of syllabic information. Thus, cortical synchronization would allow to track relevant linguistic temporal markers in the speech signal, segmenting it into syllabic-size units. Crucially, cortical oscillations have also been shown to be hierarchically nested (Figure 1.3). Gross et. al showed that, not only do they synchronize to the speech input, but also, they are coupled among frequencies, i.e., brain-brain synchronization (Gross et al., 2013).. In particular, in auditory cortex, the phase of theta oscillations was shown to modulate the amplitude of gamma oscillations This could allow for the binding and integration of linguistic information at the syllabic and phonemic level (Giraud & Poeppel, 2012; Hyafil et al., 2015).

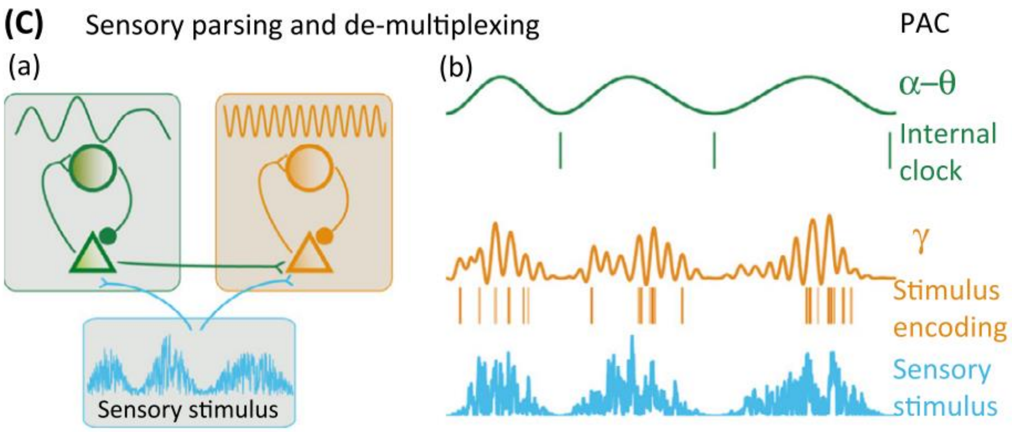


Figure 1.3. Sensory tracking and nesting of cortical oscillations. (a) slow and fast cortical oscillations synchronize to a sensory stimulus, e.g., speech, as well as to each other (b) the phase of the slow oscillation modulates the amplitude of the fast oscillation, connecting the timing of the sensory stimulus (when) to the decoding of the fast oscillation (what). PAC: phase-amplitude coupling [From Hyafil, Giraud, Fontolan, & Gutkin, 2015].

Taken together, these lines of evidence support the role that brain-speech and brain-brain synchronization have in extracting and integrating linguistic information at different grain sizes from the speech signal, and thus in the development of phonological representations. In the next section we focus on how speech segmentation through cortical oscillations relates to reading acquisition.

2.2. Rhythms in reading

If brain-speech synchronization underlies speech segmentation and the development of phonological representations, then poor brain-speech synchronization could explain reading difficulties. This is the rationale behind the Temporal sampling framework (TSF) (Goswami, 2011, 2018).

The TSF proposed that poor synchronization to speech could explain phonological deficits at the phoneme and syllabic/prosodic levels[1] (Figure 1.4). The term *poor synchronization* in this scenario refers to a decreased sensitivity to rise time in the speech envelope. As explained above, rise time is defined as the time taken to reach the peak in a speech edge, which corresponds to the occurrence of syllables. While both stressed and unstressed syllables are marked by speech edges, rise time distinguishes stressed from unstressed syllables. In stressed

---

[1] TSF also made predictions on the role that asymmetric hemispheric processing plays in phonological deficits in dyslexics. We do not include these in the description since they are not directly related to our testing paradigm.

syllables, rise time is slower. Moreover, stressed syllables underly prosodic perception in speech, occurring in the delta rate (approximately two per second). Thus, reduced sensitivity to rise time could result in poor temporal alignment of cortical oscillations to speech (i.e., synchronization). Accordingly, at the behavioural level, Leong et al., (2011) showed that dyslexics readers perform poorly in stress detection tasks. A first task involved a discrimination task for tones varying in rise time. A second task involved making same-different judgments in pairs of words differing only in their stress pattern (e.g., DIfficulty vs. diFFIculty). This was evidence that sensitivity to stress (i.e., rhythmic sensitivity) might underly the phonological deficit in reading difficulties. To explain how *phonemic* deficits might arise as a consequence of poor synchronization at slower frequencies (delta/theta), the TSF pointed to the hierarchical nesting of brain and speech rhythms. Poor synchronization at low frequencies could indirectly affect gamma synchronization, resulting in poor phonemic representations even when fast gamma oscillations are not directly impaired.

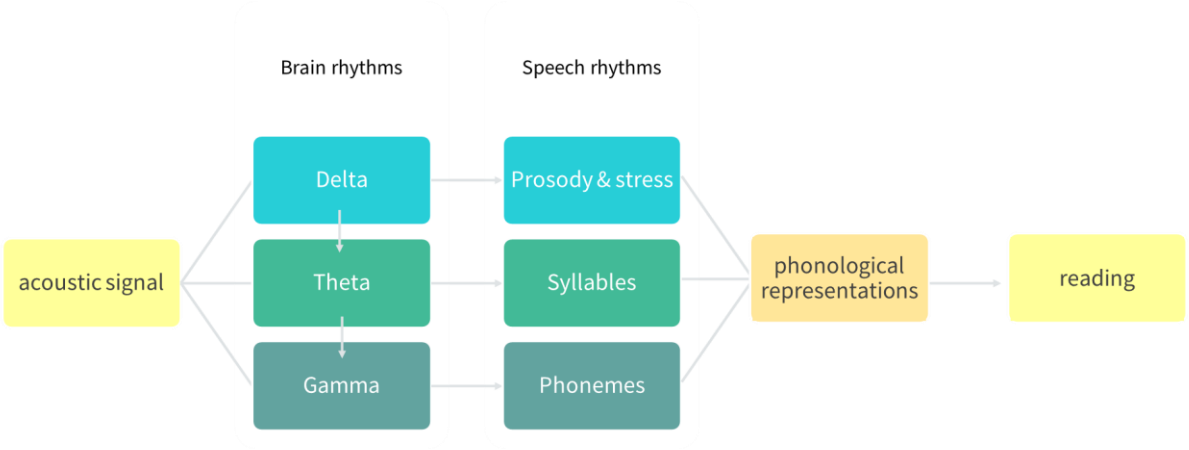


Figure 1.4. Schematic illustration of brain and speech rhythms, and its involvement in reading acquisition

Evidence in support for the temporal sampling framework at the neural level has stem mainly from studies comparing brain-speech synchronization in dyslexic versus typical readers. However, studies have varied greatly in their paradigms and methods Table 1.1. First, since the hypothesis is centred on sensitivity to *amplitude modulations* (i.e., stress), studies have used both speech and non-speech stimuli. The latter generally consists of amplitude modulated white noise. This has the advantage of isolating the critical temporal feature of the stimulus while excluding any linguistic contribution that could aid in synchronization. However, speech and non-speech stimuli differ in additional dimensions other than the linguistic content. Speech rhythm is quasi-regular, while the amplitude modulated noise commonly used is regular. An intermediate approach has used noise vocoded speech, which preserved the quasi regularity of speech but excludes its linguistic content. Second, several different analytic approaches have also been used, mainly in two dimensions, focusing on how the *phase* or the *amplitude/power* of the ongoing cortical oscillation is modulated by the auditory stimulus. While, as explained above, neural evidence points to phase alignment as the underlying mechanism for brain-speech synchronization, when the stimulus is regular, phase alignment necessarily results in increased power. The opposite does not hold though, for quasi-rhythmic stimuli, phase alignment does not necessarily result in increased power. Even within studies focusing on phase or power, there are methodological differences in how these are operationalized. Phase alignment has been tested through coherence or phase-locking values; synchronization in power/amplitude has been tested through signal to noise ratios or cross-correlations between the brain and speech signals. A last difference, though mostly terminological, stems from the use of the terms *entrainment* and *synchronization*. Entrainment implies that one signal becomes coupled to another signal. Synchronization is a more general term in which two signals are aligned. The main difference stems from one signal *driving* another, versus each of them oscillating —in synchrony, but— independently. Since the term entrainment is often used without actually proving entrainment, we will prefer the term *synchronization* to refer to both entrainment in the narrow and in the broad sense (Obleser & Kayser, 2019).

Table 1.1 Summary of studies analysing neural synchronization and dyslexia. From Lizarazu et al., 2021

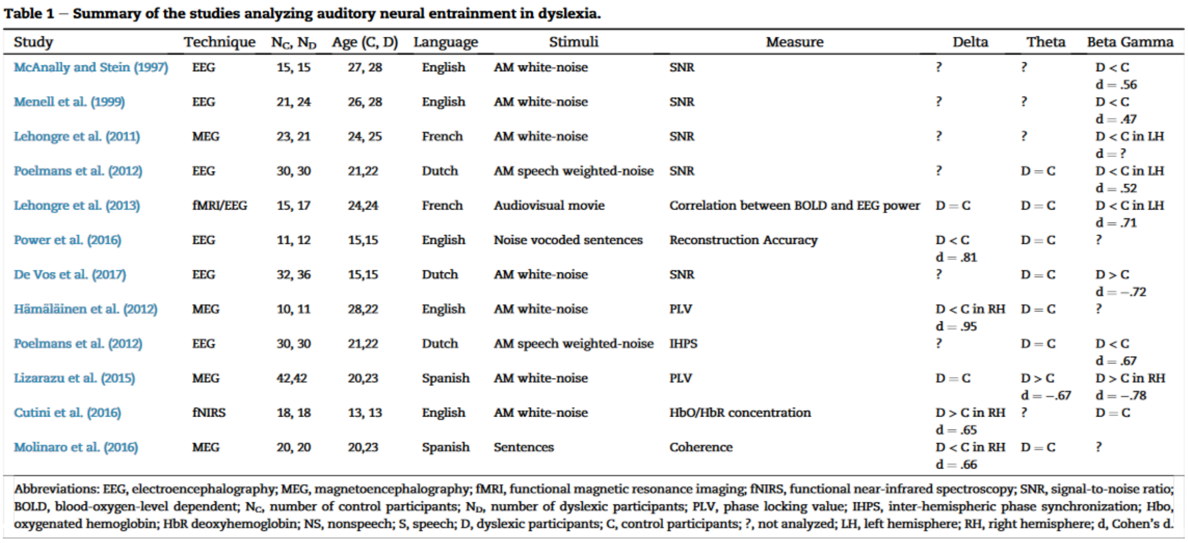


Table 1 – Summary of the studies analyzing auditory neural entrainment in dyslexia.

Study	Technique	N_C, N_D	Age (C, D)	Language	Stimuli	Measure	Delta	Theta	Beta Gamma
McAnally and Stein (1997)	EEG	15, 15	27, 28	English	AM white-noise	SNR	?	?	D < C d = .56
Menell et al. (1999)	EEG	21, 24	26, 28	English	AM white-noise	SNR	?	?	D < C d = .47
Lehongre et al. (2011)	MEG	23, 21	24, 25	French	AM white-noise	SNR	?	?	D < C in LH d = ?
Poelmans et al. (2012)	EEG	30, 30	21,22	Dutch	AM speech weighted-noise	SNR	?	D = C	D < C in LH d = .52
Lehongre et al. (2013)	fMRI/EEG	15, 17	24,24	French	Audiovisual movie	Correlation between BOLD and EEG power	D = C	D = C	D < C in LH d = .71
Power et al. (2016)	EEG	11, 12	15,15	English	Noise vocoded sentences	Reconstruction Accuracy	D < C d = .81	D = C	?
De Vos et al. (2017)	EEG	32, 36	15,15	Dutch	AM white-noise	SNR	?	D = C	D > C d = −.72
Hämäläinen et al. (2012)	MEG	10, 11	28,22	English	AM white-noise	PLV	D < C in RH d = .95	D = C	?
Poelmans et al. (2012)	EEG	30, 30	21,22	Dutch	AM speech weighted-noise	IHPS	?	D = C	D < C d = .67
Lizarazu et al. (2015)	MEG	42,42	20,23	Spanish	AM white-noise	PLV	D = C	D > C d = −.67	D > C in RH d = −.78
Cutini et al. (2016)	fNIRS	18, 18	13, 13	English	AM white-noise	HbO/HbR concentration	D > C in RH d = .65	?	D = C
Molinaro et al. (2016)	MEG	20, 20	20,23	Spanish	Sentences	Coherence	D < C in RH d = .66	D = C	?

Abbreviations: EEG, electroencephalography; MEG, magnetoencephalography; fMRI, functional magnetic resonance imaging; fNIRS, functional near-infrared spectroscopy; SNR, signal-to-noise ratio; BOLD, blood-oxygen-level dependent; N_C, number of control participants; N_D, number of dyslexic participants; PLV, phase locking value; IHPS, inter-hemispheric phase synchronization; Hbo, oxygenated hemoglobin; HbR deoxyhemoglobin; NS, nonspeech; S, speech; D, dyslexic participants; C, control participants; ?, not analyzed; LH, left hemisphere; RH, right hemisphere; d, Cohen's d.

In adults, studies have shown differences in neural synchronization between dyslexics and controls in delta (Hämäläinen et al., 2012), theta (Lizarazu, Lalllier, et al., 2015) and gamma synchronization (Lehongre et al., 2011; Lizarazu, Lalllier, et al., 2015). In children, differences in neural synchronization have also been observed in delta (Cutini et al., 2016; Lallier et al., 2016; Power et al., 2013, 2016), theta (Lizarazu, Lalllier, et al., 2015) and gamma (Lehongre et al., 2011). However, as described above, these studies stem from diverse languages, populations, stimuli, techniques, and methods, which makes it hard to integrate them (Table 1.1). For example, at the delta rate, one study found dyslexics showed weaker synchronization than controls (Molinaro et al., 2016), another study found dyslexics showed stronger synchronization than controls (Cutini et al., 2016), whereas a third one found no differences (Lizarazu, Lalllier, et al., 2015). Additionally, a recent study trying to replicate the overall previous findings mostly failed to do so (Lizarazu et al., 2021).

In sum, partial support for the TSF is available, but a detailed account on its mechanisms is still lacking. Moreover, a major unresolved issue, which cannot be addressed by the design in the previous described studies, is the causal relation between neural synchronization and reading. In the following section we focus on this issue.

2.3. Summary

The perception of speech rhythm arises from amplitude modulations in the speech signal, corresponding to the rate of (stressed) syllables. These acoustic landmarks serve as a cue for speech segmentation, which is achieved by the synchronization of brain and speech rhythms. Thus, rhythmic sensitivity at the cognitive and neural level arises as a strong candidate to underly the development of phonological representations and, therefore, it should be directly related to reading acquisition. However, a convergent picture of the underlying mechanisms is still missing.

3. Reciprocal effects between reading experience and preliteracy skills

Dyslexia[2] is a difficulty with reading that cannot be explained by poor sensory deficits or inadequate instruction (Protopapas, 2019). There is currently high agreement that the most frequent cause of reading difficulties in dyslexics stems from a phonological deficit (Ramus, 2004). However, the sources of the phonological deficits are much more debated. One strand of research points to sensory deficits as the main cause. Another, holds that phonological deficits itself are the proximal cause and that sensory deficits often reported in the dyslexic population are rather coincidental (White et al., 2006). In a provoking article, Falk Huettig and colleagues have questioned the causal role that phonological deficits may have on reading difficulties (Huettig et al., 2018). They suggest that many of the impairments reported in dyslexics —including the phonological deficit— may arise as a consequence rather than a cause of reading difficulties. The main argument suggests that "a substantial number of dyslexia

---

[2] In the present work we will use the term *reading difficulties* to refer to any type of reading difficulty at the early stages of reading acquisition, including dyslexia. Given the study design, we will not —and could not— distinguish reading difficulties that arise from poor instruction from those that arise from underlying inherited risk factors. For a more thorough description of this difference please refer to Protopapas (2019).

research finding are a proxy for a lack of (adequate) reading experience" (2018, p. 334). Importantly, in their terms, reading experience includes informal contact with reading, beyond the explicit "book reading" experience, such as writing on social media. Thus, they point that differences in reading experience between dyslexics and typical readers are both quantitative (as it is generally regarded) but also qualitative. Overall, they point to a central confound: that reading experience has reciprocal effects on many of the proposed causal factors for dyslexia, including phonological awareness (Castles & Coltheart, 2004), verbal short-term memory (Smalle et al., 2019), and rapid automatized naming (Araújo et al., 2019). Finally, they claim that many of the impairments observed in dyslexics are also observed in illiterates, suggesting that these impairments do not play a causal role in dyslexia. They close their article with four recommendations for future dyslexia research, one of which is central to the present work: *large scale longitudinal studies* (see also Goswami, 2003 and; Goswami, Power, Lallier, & Facoetti, 2014 for a similar argument). These fulfil the causality criterion of *temporal precedence* and —when focusing on prereaders at the beginning of the study— and get closer at controlling for differences in reading experience.

4. The present work

In the previous sections we have succinctly described some of the current understanding of the cognitive and neural basis of reading acquisition and pointed to some of its unresolved issues: i) how orthographic consistency modulates the role of preliteracy skills in reading acquisition, ii) how brain-speech synchronization underlies phonological development, and iii) how differences in reading experience obscure the causal role of preliteracy skills in reading acquisition.

In the next chapters we present four studies trying to address these issues. Study One: A universal screener for reading difficulties, aims at developing a digital screener to identify kindergarten children at-risk of developing reading difficulties. This is a paramount endeavour in order to set in place timely effective interventions. For it to be applied widely, it needs to be feasible and cost-effective while maintaining high classification accuracy. We present a digital screening tool and test how it fulfils these conditions. While this does not directly address any of the issues enumerated above, it addresses a more pragmatic problem: how to timely identify and support children at risk of reading difficulties. Study Two: Cognitive substrates of reading acquisition, aims at unravelling the role that preliteracy skills play in reading acquisition when examining it in a transparent language. This contributes to understanding how universal reading acquisition is, and which specific differences arise across languages depending on the characteristics of their orthography. Study Three: Dissecting the contribution of rhythmic sensitivity, addresses how rhythmic skills relate to reading acquisition at the cognitive level. In particular, it focuses on the intervening role that phonological processing plays in the rhythm-reading link. Finally, Study Four: Neural synchronization and reading acquisition, addresses the rhythm-reading relation at the neural level. It examines, through an electroencephalographic study, whether neural synchronization to rhythmic stimuli in prereaders can explain future reading acquisition.

Overall, the present doctoral dissertation is framed in a translational framework bridging neuroscience, cognitive and educational research (Dresler et al., 2018). One major contribution across studies stems from its longitudinal design, providing supporting evidence for a causal role for the proposed cognitive and neural substrates of reading acquisition. Secondly, it offers evidence from reading acquisition in a transparent orthography, which is infrequent. Third, it addresses these issues in an ecological context by assessing children in the school setting, through our comprehensive digital tool. Such an approach made it possible to test a large sample of approximately 600 kindergarten children and to control for a vast number of covariates, increasing its power and reducing the risk for confounds, respectively. In sum, we present a longitudinal study on the cognitive and neural substrates of reading acquisition in a transparent orthography, spanning digital screening in the school context to neural synchronization at the laboratory.

Chapter 2

Methods

1. Sample and design

Sampling comprised 26 public schools in Montevideo (Uruguay). All schools were above the fourth quintile in socioeconomic status (Q4 = 9 schools, Q5 = 17 schools), according to the public school system rating (Administración Nacional de Educación Pública, ANEP). Schools were either part-time or full-time. All children attending K5 level at Time 1 (821 children) were invited to take part in the study. Only those whose parents signed the consent form finally took part. Sample size at Time 1 included 616 (75%) children. At Time 2, 397 (64.4 %) out of the original 616 children continued in the study. According to the data available in the public-school system database (GURI), 76% of the children continued in G1 at the same school where they had attended K5, 5% moved to private schools and 13 % switched between public schools. At Time 2 one of the schools dropped out of the study for scheduling reasons (2.5% of children). The remaining 2.5 % could not be tracked (most of them due to a mismatch between their ID number in our database and the one in GURI). At Time 3, all children that had taken part at Time 1 or Time 2 and that were still attending any one of the 26 participating schools were invited to continue the study, except for 5 schools that could not continue for scheduling reasons (92 children). At Time 3, 250 children continued in the study (62.9 % of Time 2 sample, 40.5 % of Time 1 sample). We do not have access to the mobility occurring between Time 1 and Time 2 thus we cannot describe the reasons for the dropout.

Time 1 data collection took place in the second trimester of the school year, between June and August 2016; Time 2 and Time 3 data collection took place in the last trimester of the school year, between October and December 2017 and 2018 (in Uruguay the academic year starts in March and ends in December).

Children were assessed at their School, in groups of 4 to 5. Each child was assessed in 4-5 sessions, approximately 20 minutes each in Time 1 and Time 2, and 1 session of 20 minutes at Time 3 (only reading measures were included at this timepoint). Two research assistants monitored task performance and were available to clarify instructions on demand.

2. Measures

All tasks were presented using a tablet-based App -*Lexiland*- developed by the research team (Figure 2.1). In order to increase children's motivation and engagement in autonomous play, tasks where embedded in a videogame-like ludic narrative, with a main character and rewards for task completion. All tasks consisted of two to three example trials, four to five practice trials with feedback, followed by test trials without feedback. Effort was made to avoid the need to obtain verbal responses, in order to automatize data collection and processing. Thus, verbal responses were replaced by multiple choice items when possible (except for the Reading and RAN tasks). Instructions and auditory stimuli were pre-recorded and presented via headphones. Response times and errors were recorded in all tasks.

Figure 2.1 Lexiland videogame screenshots. Left to right, top to bottom: Segmentation, Blending, Onset matching and Rhyme, Letter Knowledge, RAN, Vocabulary, verbal Short-term memory, non-verbal Short-term memory, IQ.

2.1. Phonological awareness (PA)

Phonological awareness was assessed through four tasks: segmentation, blending, onset matching and rhyme. For each task, two separate subtasks at the syllable and phoneme levels were presented (except for rhyming).

- Segmentation (22 syllabic items, 28 phonemic items): a word was presented aurally together with a picture of it, children were asked to segment a word in either syllables or phonemes. In order to avoid verbal responses, together with the picture of the word, illustrations of dices corresponding to number two to four for syllables, and three to five for phonemes appeared in the screen. The answer was given by tapping on the dice corresponding to the number of syllables or phonemes in the word. Within each grain size, items ranged from two to four syllables, and three to five phonemes. Within each length, approximately half of the items began with CV syllables, and half with CCV syllables.

- Blending (18 syllabic items, 16 phonemic items): children were asked to blend aurally presented syllabic or phonemic segments into a word. The answer was given by selecting one out of four pictures presenting the target word and three distractors (one semantically related, one phonologically related and one unrelated). Within each grain size, items ranged from two to four syllables, and four to six phonemes. Within each length, approximately half of the items began with CV syllables, and half with CCV syllables.
- Onset matching (27 syllabic items, 32 phonemic items) and rhyme (10 word items, 10 pseudowords items): children heard pairs of words (rhyme also included pseudowords) and saw pictures for each of them (except for pseudowords). They had to answer whether both words started with the same syllable or phoneme (isolation) or rhymed (rhyme). The answer was given by tapping on a tick or a cross on the screen. For onset matching, within each grain size, items ranged from two to three syllables, and four to six phonemes. Within each length, approximately half of the items began with CV syllables, and half with CCV syllables. For rhyme, all items had three syllables and a CV syllable structure.

2.2. Letter knowledge (LK)

Letter knowledge was assessed separately for letter name and letter sound. In each subtasks the name or sound of each letter was presented aurally and children were asked to choose the correct visual letter among three options: the target, a visually similar distractor (Boles & Clifford, 1989) and an unrelated distractor. There were 22 items of each type [for a total of 44].

2.3. Rapid automatized naming (RAN)

Children were presented with an array of 5 items repeated 6 times each and were asked to name them as fast and as accurately as possible. Items were either objects (*gato, jugo, mano, silla, queso* [cat, juice, hand, chair, cheese, respectively]), colours (*azul, negro, rojo, verde, blanco* [blue, black, red, green, white]), numbers (4, 5, 7, 8, 9) or capital letters (F, M, N, S, R). Number of errors and total time were recorded. All children were presented with the 4 subtasks. All subtasks were preceded by a familiarization phase where they were asked to name each item separately to ensure that they knew its name.

2.4. Vocabulary (VOC)

At Time 1, Receptive vocabulary was measured through the BEST test (De Bruin et al., 2017). Given that the accuracy results at Time 1 suggested ceiling effects, at Time 2 a digital version of the Peabody Picture Vocabulary Test (Dunn et al., 2006) was added. The procedure was the same as in paper, except that the response was given by tapping on the screen.

2.5. Short-term memory (STM)

2.5.1. Verbal

Verbal short-term memory was assessed through an adaptation of the task described in Martinez Perez, Majerus, & Poncelet (2012). Monosyllabic words were presented aurally (*sol, pan, tren, rey, flor, pez* [sun, bread, train, king, flower, fish]) followed by images corresponding to the words heard. Children were asked to order the images according to the order of aural presentation. The sequence ranged from 2 to 6 items.

2.5.2. Non-verbal short-term memory

Visuo-spatial short-term memory was assessed through an adaptation of the Corsi Block Tapping task (Corsi, 1972). Blocks were replaced by pictures of pigs to make it more attractive for children. Sequences ranged from 2 to 8 elements. Testing was interrupted if 3 errors were made on 4 consecutive trials of the same length.

2.6. Nonverbal IQ (IQ)

Nonverbal IQ was measured using the Matrix Reasoning subtest of the Spanish version of the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 2001).

2.7. Reading

2.7.1. Decoding (Time 1 & 2)

At Time 1, a list of 15 frequent words and 15 pseudowords was presented in paper; children were asked to read them aloud. At this point children were not expected to read given the guidelines of the Education System in Uruguay for K5. Number of errors was recorded. At Time 2 the reading assessment included two subtasks: i) decoding of a list of 30 words and 30 pseudowords presented digitally, one word per screen; ii) word and pseudoword decoding of the PROLEC-R battery (Cuetos et al., 2007), in paper, which consists of 80 items.

2.7.2. Fluency (Time 2 & 3)

Fluency was assessed via a two-minute reading test that consisted of reading as fast and as accurately as possible a meaningless text of 278 words in 2 minutes. The text was presented in paper. Number of words read, and number of errors were recorded.

2.7.3. Comprehension (Time 2 & 3)

Reading comprehension was assessed through the sentence comprehension subtask of the PROLEC-R battery (Cuetos et al., 2007). The task consists of 17 items of increasing complexity. The first 3 items consist of reading and performing an action (i.e., "tap on the table three times", the following six items consist of reading and completing a drawing (i.e., "draw three apples in the tree") and the last 7 items consist in reading and choosing one out of 4 pictures (i.e., "the horse is smaller than the elephant"). Thus, the first 9 items were presented in paper, and the last 7 items were presented digitally. Sentences were written in uppercase format.

3. Psychometric properties of the phonological awareness tasks

Given the central role that phonological awareness plays in reading acquisition and considering that we have employed an atypical form of measurement, we examined the psychometric properties of the phonological awareness tasks. Given its extension, it is included in Appendix A.

4. Reading outcomes across studies

Different reading outcomes are used across Studies, responding to different research questions. Study 1 uses a composite score of decoding, fluency, and comprehension, study 2 uses decoding and fluency separately, study 3 uses efficiency, and study 4 uses decoding only. At first glance, these decisions might seem incongruent, and they would be, if our aim were to directly compare results across studies. However, this is not the general aim. The selection of the reading outcome in each study responds to the theoretical questions addressed. The first study aims at developing a screener tool for classification of children into at-risk (poor readers) and not-at risk (typical readers) of developing reading difficulties. Reading here is considered in its broader more comprehensive sense. Not only do we want to identify children with poor decoding skills, but we also want to identify low achievers in fluency and/or comprehension. Naturally, these three reading components are highly correlated, but they are also independent. Combining the three reading variables allows us to operationalize reading in its broader sense and search for at-risk children irrespective of their source. The second study aims at understanding the cognitive basis of reading acquisition in a transparent orthography. It is expected, by hypothesis, that these might differ across reading outcomes, and thus decoding and fluency are modelled separately. Comprehension is not included since measured preliteracy skills are focused on the acquisition of decoding and fluency skills, and less so on comprehension (which is more dependent on oral language and metacognitive skills (Muter et al., 2004)). For this same reason, we do not focus Study One in trying to understand the contribution of each preliteracy skill in future reading achievement. Explaining all three reading components at the same time would generate more confusion than shedding light on the cognitive basis of reading acquisition. Study Three aims at understanding how reading relates to rhythmic sensitivity, and the intervening role of phonological processing in this measure. Thus, the reading outcome is decoding. We compute a ratio of decoding accuracy over speed (termed efficiency) in order to account for both aspects of the reading process. Finally, Study Four includes only decoding since, by hypothesis, we expect neural synchronization to auditory stimulus to impact mainly on the development of phonological representations and thus decoding.

5. Item response theory

Item response theory (IRT) allows, among other advantages, to estimate a subject latent trait from observed responses. In the most common model, the 2-parameter model, each item is modelled as a logistic curve defined by 2 parameters: discrimination and difficulty. These correspond to *a* and *b* respectively in the following equation:

$$P(x = 1 \mid \theta) = \frac{1}{1 + e^{-a(\theta+b)}}$$

In a simpler form, the Rasch model, the *a* parameter (discrimination) is fixed constant, and items are allowed to vary only in their difficulty levels. Difficulty can be interpreted as the ability level where the probability of a correct answer is 0.5. Difficulty values range approximately between -3 and 3. Besides the item parameters, IRT allows to estimate a subject parameter. As seen from the equation, the probability of a correct response is expressed as a function of the item parameters and the *theta* (θ) parameter, an estimate of a subject's latent ability. For this purpose, item and subject parameters are estimated iteratively with Marginal Maximum Likelihood Estimation (MMLE), until reaching a model compatible with the observed data. Further details of estimation of model parameters can be found on Rizopolous (2006) and Baker (2001).

Since IRT allows to estimate difficulty and discrimination levels for each item, it can be used to refine an assessment tool by discarding items with poor psychometric properties. Additionally, since it also allows to estimate subject parameters, it can be used to obtain latent trait scores for individual subjects. IRT has been previously used to study the PA construct in English (Schatschneider et al., 1999) and Dutch (Vloedgraven & Verhoeven, 2007, 2009). In Chapter 4 we used this approach in order to test the contribution of phonological awareness to reading acquisition.

6. Linear mixed effects models

Unless otherwise noted, all mixed effects regression models were computed through the lme4 package (Bates et al., 2015) in the R software (R Core Team, 2018). Overall, linear mixed effects models were used to account for the multi-level structure of the data, with children nested within Schools. Thus, School was included as a random intercept in most models. By doing so, the model accounts for the variance explained by the School effect, generally improving model fit.

Chapter 3

Study One: A universal screener for reading difficulties

1. Introduction

At first glance, a child facing difficulties with learning to read in first grade is much like the case of the tortoise and the hare: he or she is just taking a bit more time than their peers, but will catch up eventually, he or she just needs more time. This same logic holds for the second grade. In third grade, if he is still struggling, then he is referred to a specialist who, in many cases, will make a Dyslexia diagnosis. Only then will the child be directed towards a personalized remedial treatment. This is the current protocol in the United States and Uruguay alike (Seidenberg, 2017, Chapter 11), and is likely the case in many other countries, as well. While the reasoning behind this sounds intuitive, it is however inaccurate. Most children showing signs of reading difficulties in the very early stages of learning to read, will persevere in their difficulties if no intervention is set in place (Ozernov-Palchik & Gaab, 2016). Moreover, it has dire consequences for children undergoing such difficulties. Several studies show that children with reading difficulties face exclusion from the educational system, limitations in their socioemotional development, and higher rates of depression and anxiety (Arnold et al., 2005; Sprenger-Charolles et al., 2011). Moreover, struggling readers accumulate less reading experience than their peers, thus acquiring less vocabulary, in a downwards spiral known as a “Mathew effect” where the rich get richer and the poor get poorer (Stanovich, 1986). In Uruguay, a report from 2016 based on PISA scores, showed that only 53% of 15-year-olds attain minimal competences in reading, which is also a strong predictor of dropout risk. Among children that do attain sufficient reading competence, 90% will finish high school, while among children that do not attain sufficient reading competence, only 17% will (Cardozo, 2016; INEEd, 2016).

An alternative to this wait-for-failure approach is prevention. In the past four decades, research in cognitive science has found a set of skills that develop before reading instruction, referred to as preliteracy skills, that are strong predictors of future reading difficulties. These preliteracy skills include phonological awareness (PA), letter knowledge (LK), and rapid automatized naming (RAN) (Boets et al., 2007; Lyytinen et al., 2006a; Muter et al., 2004; Schatschneider et al., 2004). PA refers to the ability to identify and manipulate the sound structure of the oral language and is usually measured in tasks that require, for example, segmenting words into their constituent syllables or phonemes. LK is the ability to map letter names or sounds to their corresponding written representations. The RAN task measures naming speed and lexical access by presenting a grid of objects, colours, letters, or numbers that the participant has to name as quickly and accurately as possible. Thus, by assessing these preliteracy skills in kindergarten, it is possible to identify children at risk of developing reading difficulties early on, and thus profit from early intervention.

In medicine, the term *screening* refers to this approach, namely, testing for risk markers of a future condition, typically in order to provide early intervention. Screening differs from diagnosis in that the condition has not yet developed, and thus cannot be diagnosed. The goal of screening is to identify individuals who are likely to develop a certain condition in the future and, when possible, quantify the probability that they will develop the condition. Screening can target a particular group of individuals—for example those with higher risk due to a genetic predisposition—or it can be universal, that is, targeting all individuals in a certain population. For example, universal screening for hearing loss is performed on every newborn in the United States, Uruguay, and Spain (Calonge, 2008; Ministerio de Salud, 2017). For universal screening to be effective, a set of criteria need to be met. In the following section, we detail these criteria.

1.1. Desirable features of a universal literacy screener

A universal screener needs to be timely. Despite the fact that remedial interventions are more effective the earlier they begin, it is common practice for diagnosis and referral to wait until children show evident signs of underperforming compared to their peers, which generally occurs around third grade (Ozernov-Palchik & Gaab, 2016; Wanzek & Vaughn, 2007). At this point, children have struggled with reading for two or three years and have accumulated less reading experience and developed a negative attitude towards literacy overall (Stanovich, 1986). Thus, not only the optimal window of opportunity has been lost, but also novel cascading negative effects need to be overcome.

A universal screener needs to be feasible and cost-effective for large samples. While many research studies have shown the predictive validity of preliteracy skills as longitudinal predictors of future reading success (Andrade et al., 2015; Catts et al., 2009; Furnes & Samuelsson, 2010; Lonigan et al., 2000; Moll et al., 2014; Muter et al., 2004; Peng et al., 2019; Puolakanaho et al., 2007; Thompson et al., 2015), testing preliteracy skills in an ecological environment and in a large sample is not trivial. In a research context, preliteracy skills are usually assessed individually by a trained researcher or research assistant, with sample sizes in the order of tens to a few hundreds, and lately—but rarely—closer to one thousand (see for example Ozernov-Palchik et al., 2017). However, if screening is to be applied to hundreds of thousands of children, the individual approach is hard to sustain, especially in developing countries. The cost-effectiveness of screening can be dramatically improved through digital screening, made possible by recent technological developments. Digital screening has many potential advantages. First, it allows the assessment tasks to be “gamified”, increasing children’s motivation and engagement, and making it possible for collective self-assessment. Second, responses can be recorded and automatically processed, without the need for trained staff. Third, data collection could be ongoing and continuously updated, such that local up-to-date norms can be calculated. This last point is a long-standing issue for the more precise identification of risk profiles, as norms vary by population and literacy stage. All these advantages combined can significantly decrease the cost of early screening, making wide-spread implementation possible.

A universal screener needs to have high sensitivity and specificity. In the signal detection framework, classification errors can arise from two sources. One refers to identifying the signal as present, when in fact it is not. The complementary case is identifying the signal as absent, when in fact it is present. These two types of errors are called false positive (FP) and false negative (FN), respectively. While in some cases the two types of errors have equivalent consequences, in others they have different costs. For example, if the classifier is a fire detector, it is much less costly to ring the alarm when there is no fire (FP), than to not ring the alarm when there is a fire (FN). In other cases, this is not the case. When the classifier refers to detecting a condition, an FP is costly for the public system—whether its health or education—as it is providing treatment for a condition that is absent, and also for the persons and families who are misguided into thinking they are subject to a condition they do not have. The opposite of a classification error is correct classification, which can occur in two ways: saying the signal is present when it is indeed present (true positive, TP) or saying the signal is absent when it is indeed absent (true negative, TN). The ratio of TP to the total number of positive cases—namely, the cases in which the condition is in fact present—is called *sensitivity*, and the ratio of TN to the total number of negative cases—namely, the number of cases in which the condition is absent—is called *specificity*. When creating tests, there is always a trade-off between sensitivity and specificity. For example, a classifier trying to identify children at-risk of developing reading difficulties could classify all children as at-risk, thus having 100% sensitivity, but erroneously classify all children who are not at-risk as at-risk, thus having 0% specificity. Previous studies in reading have obtained the best sensitivity and specificity when behavioural predictors are combined with brain measures such as EEG or fMRI, reaching up to 90% sensitivity and 80% specificity (Hoeft et al., 2011; Molfese, 2000). Unfortunately, brain measures greatly increase the cost of screening, making it unfeasible for large populations. Another approach to improving sensitivity and specificity has been to include response to intervention (RTI) in the screening process (Vellutino et al., 2008). That is, including individual gains in preliteracy or literacy skills during in group intervention to predict future reading gains. This approach yielded 95% sensitivity and specificity levels in a sample of approximately 120 children when RTI measures where included, and 68% sensitivity and 72% specificity when only initial screening scores were included in a sample of approximately 400 children. Thus, when only single-assessment behavioural measures are used, sensitivity and specificity are generally lower. For example, in the Jyvaskyla Longitudinal Study of Dyslexia in Finnish, a 90% sensitivity was obtained with 65% specificity when LK, RAN, and familial risk of dyslexia were assessed at 5.5 years of age in a sample of 200 children (Puolakanaho et al., 2007). Equivalent levels were obtained in a study in English with 260 children (Thompson et al., 2015).

A universal screener needs to be unbiased. The sample used to build a prediction model of reading difficulties should be representative of the larger population, so that its results can be generalized without bias. Many of the aforementioned studies based their models on samples with a disproportionately high percentage of children at high-risk of developing dyslexia, either because of genetic risk or prior screening (Puolakanaho et al., 2007; Thompson et al., 2015; Vellutino et al., 2008). Naturally, this is an appropriate approach in longitudinal studies focused on advancing our understanding of the cognitive underpinnings of reading difficulties, which was the aim of these studies. However, this becomes a limitation when trying to generalize the findings to the larger population.

1.2. The present study

In sum, while great efforts have been made in specifying the desirable features of a universal screener, and many studies have addressed a wide range of them, rarely have all these requirements been met in a single study. A search in the American Institute for Research National Center on Intensive Intervention database for screening tools assessing reading in English with these features—namely, kindergarten, free, digital, group or self-administered—showed no results (https://charts.intensiveintervention.org/ascreening). In the present work, we developed a universal screener that is not only cost-effective, non-biased, and comprehensive but also short enough and feasible for school settings. For this purpose, we developed a game-like digital App, which we named *Lexiland*, targeted at children attending K5, which can be self-administered in a school setting. In order to assess this App's predictive longitudinal validity, an initial sample of 600 children was followed for three years and assessed at three time-points: mid-term K5, end of first grade, and end of second grade. Children were predicted as *poor readers* or *typical readers* based on the combination of measured cognitive and demographic variables. The screener attained high classification accuracy for first and second grade reader status.

2. Methods

See Chapter 2 for a broad description of sample composition and measured cognitive variables. A brief overview is provided in the following section.

The initial sample was composed of children that completed Time 1 (K5) and Time 2 (G1) assessments (n = 388). Cognitive variables used in the present chapter include phonological awareness (PA), letter knowledge (LK), rapid automatized naming (RAN), verbal and non-verbal short-term memory (vSTM and nvSTM), vocabulary (VOC), intelligence quotient (IQ), and reading (decoding, fluency, and comprehension). Data for reading in G2 was also included for the sample of children that were still in the study during G2. Demographic variables included Age, Gender, and Socioeconomic status (SES). SES was defined as the highest achieved level of maternal education. It was treated as an ordinal variable with 3 levels: Low—Unfinished high school or less (n = 158)—, Middle—Completed high school (n = 64)—, and High—completed a Bachelor's degree or more (n = 91). Data was missing for 75 children (19.3 % of the sample), which were excluded from further analyses.

2.1. Model specification

Two logistic regression models were fit to the data. First, a *full model* with all cognitive and demographic variables was computed. Second, a *reduced model* was fit in order to reduce the number of predictor variables. Only variables with significant contributions to discriminating reader status at the 0.05 level were retained.

The significance of model coefficients was tested under a Wald II Chi Square test (i.e., the contribution of each variable is tested above and beyond all other variables in the model), and nested models were compared through likelihood ratio tests. Model fit was estimated through Nagelkerke pseudo R^2 theoretical method built into the *MuMin* package (Bartoń, 2019)

2.2. Cross validation

In order to test how well the fitted models will perform in a new sample of children, cross-validation was performed on the reduced model. Cross validation improves the generalizability of the model by training it with one sample and testing it on a new sample of unseen data. For G1 reading scores, the model was trained on a random sample of 70% of the data, and classification accuracy was tested on the remaining 30%. This split yielded a sample size of approximately 100 children in the test set, where approximately 16 children were expected to belong to the poor readers' group. A larger split (such as an 80/20 split) would decrease the number of children expected to belong to the poor readers' category and therefore increase the chances of finding convergence issues during model fit. The procedure was repeated 1000 times in order to account for the random sampling in the cross-validation process. Next, in order to test the stability of model predictions, the model built with G1 scores was used to predict unseen G2 scores.

Model performance was assessed using ROC curves, area under the curve (AUC), and the specificity levels obtained for 90% and 80% sensitivity. ROC curves represent the ratio between true positive rate (TPR or sensitivity) and false positive rate (FPR or 1 - specificity) in any binary classification model, for different cut-off thresholds (Table 3.1). The default threshold in a binary classification model is 0.5, meaning that if the predicted probability of a single case (i.e., child) is above 0.5, it is labelled as positive (in this case, poor readers), otherwise it is labelled as negative (in this case, typically-reading). The sensitivity and specificity trade-off can be modified by changing the threshold cut-off in the binary classification model. When classes or groups are balanced (that is, when it is equally likely to belong to the at-risk or to the not at-risk class), a 0.5 threshold is appropriate. However, when classes are unbalanced, as is the case with reader status, other thresholds might produce better performance. Since the present data is unbalanced, we will focus the presentation of results on the specificity levels obtained for 90% and 80% sensitivity (instead of presenting them for 0.5 and/or 0.25 threshold cut-offs which is the common practice). Area under the curve (AUC) value ranges from 0.5 to 1, where 0.5 indicates classification at chance level, and values above 0.8 are generally deemed at acceptable (Catts et al., 2009).

Table 3.1. Types of errors and successes in a binary classification model

	Condition positive (TP + FN)	Condition negative (FP + TN)
Predicted positive	True positive (TP)	False positive (FP)
Predicted negative	False negative (FN)	True negative (TN)
	Sensitivity (TPR): $\# TP / \# TP + FN$	Specificity (TNR): $\# TN/\# \mathrm{FP} + \mathrm{TN}$

Note: # total number of children in that condition; TPR: true positive rate; TNR: true negative rate

3. Results

Reader status at the end of G1 was longitudinally predicted from the cognitive and demographic variables measured in K5 using two logistic regression models (full and reduced, see Methods section). Reader status was composed of two groups: *typical readers* (n = 324) and *struggling readers* (n = 64). The full model included all of the measured cognitive and demographic variables, while the reduced model retained only the variables that significantly contributed to the prediction of reader status with at least 95% confidence. Their performance was assessed through model comparisons with likelihood ratio tests and goodness-of-fit statistics (Akaike information criterion [AIC], Bayesian information criterion [BIC], and Log likelihood). Furthermore, in order to assess the generalizability of the models, the reduced model was refitted with cross-validation. Its relative performance was compared through classification accuracy statistics (Area under the curve AUC, sensitivity, and specificity).

3.1. Reader status

Reader status was defined as the arithmetic mean of the z scores for decoding, fluency, and comprehension (correlations: decoding and fluency: r = 0.67, decoding and comprehension: r = 0.83, fluency and comprehension = 0.67, all p values < 0.001). Distribution of decoding, fluency, comprehension, and their composite scores are displayed in Figure 3.1. Tasks show a bimodal distribution, with a subset of children showing no reading skills in either decoding, fluency, or comprehension. In line with this, and in order to partition these results to create a dichotomous variable for classification, the reading composite measure was transformed into a discrete variable with two levels. Setting a threshold for classification is a non-trivial problem that has been solved in many different ways. In the reading literature, thresholds have been set at various levels including reading composites scores below the 10th or 20th percentiles as well as below 1 SD or 1.5 SD—which in a normal distribution represent the 16th and 7th percentiles, respectively (Elbro, 1996; Maurer et al., 2009; Pennington et al., 2012; Puolakanaho et al., 2007; Thompson et al., 2015). In trying to set a meaningful threshold for our sample, we decided on using the 16th percentile (a -1 z score in a normal distribution, and -1.3 z score in our bimodal distribution) since it reached a balance between strong theoretical and pragmatic motivations. On the one hand, it yielded a poor readers group with virtually no reading skills (see Table 3.2). On the other, it provided a large enough poor readers group for cross-validation purposes. Thus, children with a reading composite score below the 16th percentile (-1.3 z score) were categorized as *poor readers* (PR, n = 64), and those above that threshold as *typical readers* (TR, n = 324).

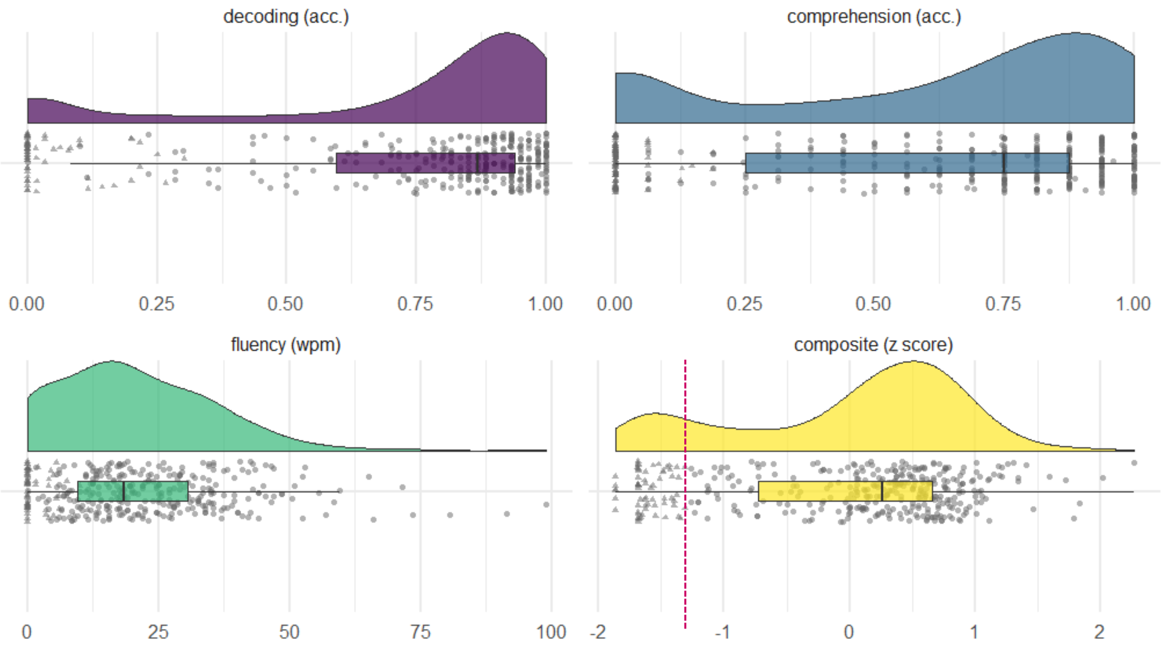


Figure 3.1. Distribution of reading scores for each reading measure. A bimodal distribution can be observed in all tasks. Triangles represent children with poor readers (PR), circles represent children typically reading (TR). Purple dashed line indicates cut-off threshold for reader status. Acc.: accuracy, wpm: words per minute.

Table 3.2 shows average reading scores by reader status. On average, TR correctly decoded 84% of the presented words, comprehended 71% of the presented sentences, and read 24 words per minute. On average, PR correctly decoded 5% of the presented words, comprehended 2% of the presented sentences, and read two words per minute. This group's average z scores on the composite reading measure were 1.6 SDs below the mean. Thus, the PR group showed virtually no reading skills by the end of first grade.

Table 3.2 Descriptive statistics for reading measures for typical readers (TR) and poor readers (PR)

	decoding *(acc.)*		comprehension *(acc.)*		fluency *(wpm)*		composite *(z)*		N
reading status	M	S	M	S	M	S	M	S	
TR	0.84	0.18	0.71	0.28	24.38	15.00	0.30	0.64	324
PR	0.05	0.08	0.02	0.05	2.13	3.10	-1.59	0.15	64

Note: TR: typical readers, PR: poor readers, M: mean, S: standard deviation, acc: accuracy, wpm: correct words per minute.

3.2. Predictor variables in the full and reduced models

The full model included all of the collected cognitive and demographic variables: Age, Gender, SES, IQ, non-verbal and verbal STM, vocabulary, RAN, letter knowledge, and phonological awareness. Among these, only SES, non-verbal STM, letter knowledge, and phonological awareness were significant predictors of reader status above and beyond all other variables in the model (Table 3.3). The Nagelkerke pseudo R2 for this model was 70%.

Table 3.3. Coefficients for the longitudinal prediction of reader status in G1 from K5 variables (full and reduced models)

model	term	estimate	std.error	statistic	p.value	conf.low	conf.high
full	(Intercept)	-3.40	0.49	-6.98	0.000	-4.46	-2.53
	Age	0.09	0.19	0.45	0.654	-0.29	0.46
	Gender (Male)	0.47	0.41	1.14	0.254	-0.33	1.28
	SES (linear)	-1.76	0.57	-3.07	0.002	-3.13	-0.78
	SES (quadratic)	-0.87	0.50	-1.74	0.082	-1.90	0.10
	IQ	-0.01	0.25	-0.06	0.953	-0.51	0.46
	verbal STM	-0.33	0.25	-1.36	0.173	-0.82	0.14
	non-verbal STM	-0.53	0.22	-2.40	0.016	-0.97	-0.10
	vocabulary	-0.17	0.20	-0.88	0.377	-0.55	0.22
	RAN	0.32	0.25	1.32	0.188	-0.16	0.81
	letter knowledge	-0.84	0.31	-2.76	0.006	-1.48	-0.27
	phonological awareness	-0.81	0.41	-1.97	0.049	-1.65	-0.04
red	Intercept	-3.20	0.42	-7.69	0.000	-4.11	-2.47
	SES (linear)	-1.76	0.57	-3.09	0.002	-3.13	-0.79
	SES (quadratic)	-0.83	0.49	-1.68	0.093	-1.86	0.12
	non-verbal STM	-0.61	0.20	-3.01	0.003	-1.02	-0.22
	letter knowledge	-1.02	0.29	-3.52	0.000	-1.64	-0.49
	phonological awareness	-1.17	0.38	-3.11	0.002	-1.95	-0.47

Note: red: reduced. SES: socio-economic-status, STM: short-term memory, RAN: rapid automatized naming. SES is an ordinal variable and thus includes a linear and a quadratic term. Conf.low: confidence interval lower bound; Conf.high: confidence interval upper bound. Confidence interval level is 95%.

With respect to the reduced model, which only included SES, non-verbal STM, letter knowledge, and phonological awareness, it did not perform significantly worse than the full model ($X^2(6) = 4.78$, $p = 0.57$). All predictor variables were significant at the 99% level. The Nagelkerke pseudo R^2 for this model was 71%. Thus, even though it reduced the number of variables that need to be measured—thus reducing assessment time—model fit was as good as the one of the full model.

In order to gain insight and interpretability from the model outcomes, the predicted probabilities of belonging to the PR group were estimated for different preliteracy skills profiles and SES levels. The following test cases were analysed: children with performance at -1 SD in either nvSTM, LK, PA, or all three, and for a child with average scores (0 z score). According to the reduced model, a child with average preliteracy skills, irrespective of SES background, has a 5.8% risk of being in the PR group. When SES is taken into account, this risk goes up to 9.2% for children from low SES homes and down to 0.0% for children from high SES homes (Figure 3.2). Low performance in any one of the preliteracy skills considered increases the risk by approximately 0.1 points for low and middle SES homes and only about 0.02 points for high SES homes. When performance is low in all of the preliteracy skills considered, the risk of being in the PR group is approximately 12% for high, 57% for middle, and 62% for low SES children.

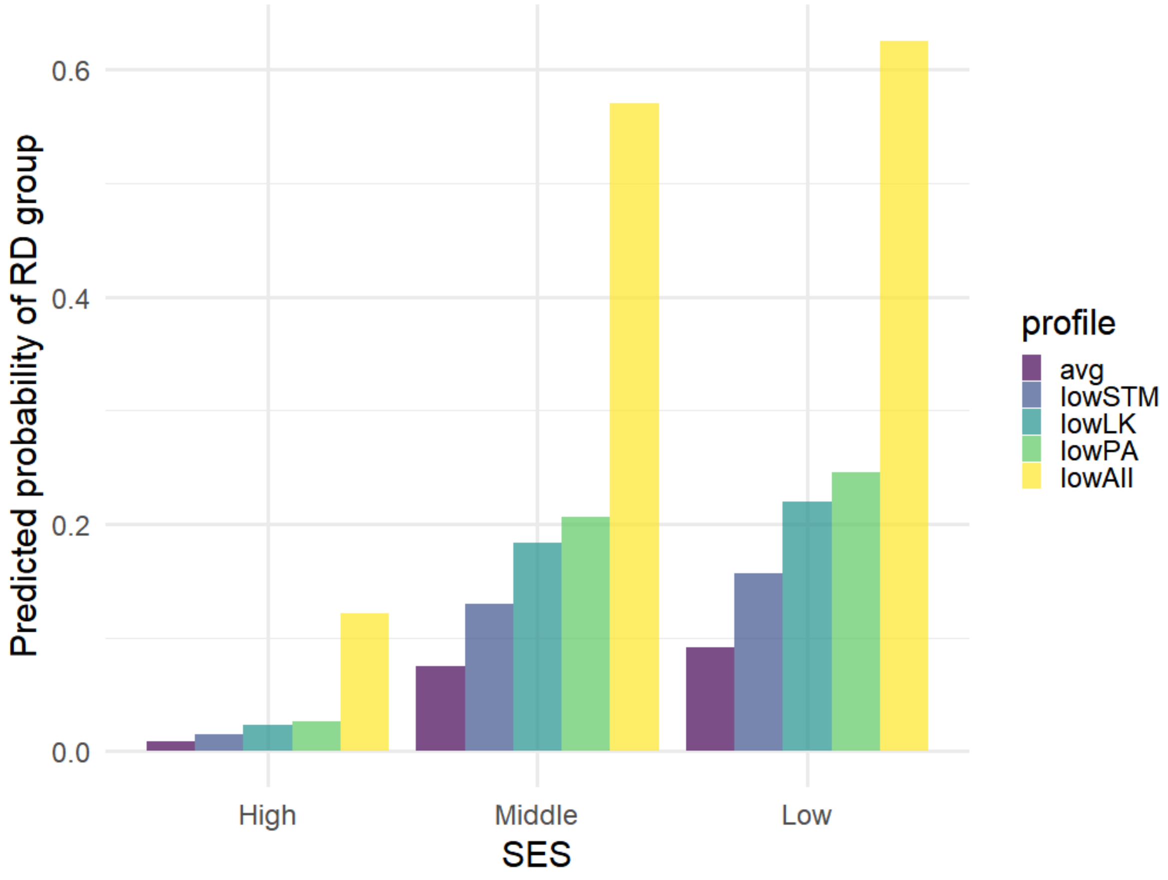


Figure 3.2. Predicted probabilities of belonging to the PR group by preliteracy skills profile and SES (reduced model)

3.3. Cross-validation performance of the reduced model

Finally, for the reduced model, cross-validation was performed in order to test its generalizability. Complete performance for each model iteration is presented in the ROC curve in Figure 3.3. The model shows high classification accuracy, with an AUC of 0.88 (min = 0.7, max = 0.97, SD = 0.04, CI.low = 0.88, CI.high = 0.89) and 76% specificity (min = 0.27, max = 0.96, SD = 0.1, CI.low = 0.75, CI.high = 0.77) for 90% sensitivity.

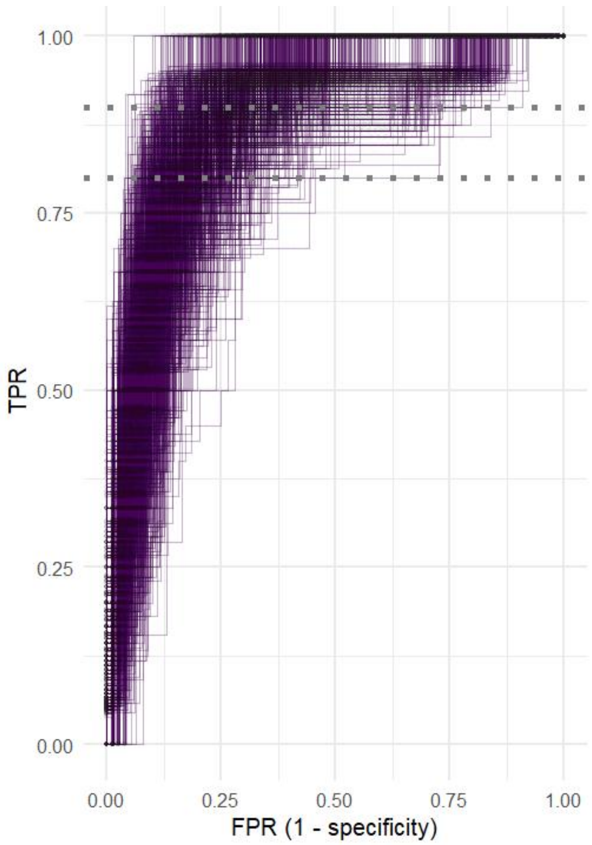


Figure 3.3. ROC curves for the reduced model. Grey dotted lines show 80% and 90% sensitivity (TPR) levels, corresponding to between 70% and 85% specificity levels.

3.4. Stability of reader status and model prediction

Two questions remain regarding the long-term trajectories of the children in the PR group. First, how stable are poor readers' trajectories. In other words, do poor readers in G1 still show reading difficulties in G2? Second, how does the classification model perform when instead of predicting reader status in G1, it is used to predict reader status in G2?

3.4.1. Reader status in G2

The proportion of children from the PR group in G1 that were still in the PR group one year later was analysed. From the 388 participants in G1, 201 children continued in the study in G2 (51%). Reading difficulties in G2 were defined following the same criteria as in G1 and contained the children in the bottom 16% of the distribution of the reading composite (the threshold for this split was -0.4 SD). Out of the 64 children in the PR group in G1, 24 remained in the sample in G2. Out of these 24, 21 (83.3%) were still in the poor readers in G2 ($X^2(1) = 90.537$, $p < 0.001$). Thus, reading difficulties showed a stable trajectory in which children with difficulties in G1 are highly likely to continue showing reading difficulties in G2.

3.4.2. Stability in model prediction

The reduced model described in Table 3.3 was used to predict reader status in G2. This model was built using SES, non-verbal STM, letter knowledge, and phonological awareness as predictor variables, and reader status in G1 as the outcome variable. Then, we asked whether the parameters obtained from that model could successfully predict reader status in G2. That is, we performed cross-validation with reader status in G2 as the test set. The model attained 61% specificity for 90% sensitivity and an AUC of 0.84.

4. Discussion

The results presented above show that it is possible to attain high classification accuracy from an early digital screener, self-administered in school. By assessing only three cognitive skills (non-verbal STM, letter knowledge, and phonological awareness), the screener correctly identified nine out of ten (90% sensitivity) kindergarteners who developed reading difficulties one year later—in G1—and nearly eight out of ten (76% specificity) who will go on to read as expected. Moreover, it also identified children who showed reading difficulties two years later—in G2—with high accuracy (90% sensitivity and 60% specificity). Thus, there is no reason for maintain a wait-for-failure approach to reading difficulties, as this entails dire consequences for the socio-emotional and professional life trajectories of children with reading difficulties (Ozernov-Palchik & Gaab, 2016).

Notably, the model showed classification accuracy levels that are equivalent to those obtained through one-on-one assessment by trained personnel (Puolakanaho et al., 2007; Thompson et al., 2015). This is in itself an accomplishment for the *Lexiland* screener and provides excellent potential as a universal screener. By using this screener with every child attending K5, it would be possible to set in place timely remediation programmes, which are known to be increasingly effective the earlier they begin (Wanzek & Vaughn, 2007). This approach's potential within small groups in schools, along with its short assessment time, is a highly valuable feature.

Socio-economic status was a significant predictor variable in all models. Unfortunately, aiming SES as a target for intervention is a much larger endeavour than focusing on preliteracy skills. Adequate teaching of letters names and/or sounds (a matter of current debate, but beyond the scope of the present chapter), together with training in phonological awareness and short-term memory should therefore be a primary concern of teachers and teachers' educational programmes. Here, again, evidence from cognitive science can inform educational practices (see for example, Sunde et al., 2019).

4.1. Educational implications

The present results show that *Lexiland* can be used as a universal screener, as it was designed to. Turning it into a public policy it is unfortunately out of our hands. Still, we believe it can be used individually by teachers interested in having a more standardized assessment of children's skills. While, of course, teachers regularly assess their students' progress, evidence shows that the correlation between teachers' assessments and standardized measures of preliteracy and literacy skills is moderate, and that teachers tend to overestimate their students' skills (Cabell et al., 2009; Martin & Shapiro, 2011). Lexiland could be used to identify at-risk children and monitor their progress throughout the school year. Moreover, an overall assessment of the whole class can aid teachers in lesson planning, for example, by targeting the letters that are not known by most of the class.

4.2. Limitations

The present study was composed of an unselected sample of children attending K5 in middle and high SES public schools in Montevideo. Despite this being an advance with respect to studies with selected samples of at-risk children, it is nonetheless not a representative sample of the entire population, and thus its generalizability is limited to children attending schools with similar demographic characteristics. It should be noted though that the sensitivity and specificity levels reported here are the result of cross-validation which, in and of itself, is a thorough test of Lexiland's model generalizability. Still, a future assessment should focus on testing a new, shorter version of the battery—possibly assessing only letter knowledge, phonological awareness, and non-verbal short-term memory—in a representative sample of children with a broader range of SES statuses and from the entire country. It is important to note, though, that the population of Montevideo represents approximately half of the country's population, so such an adjustment in sample is realistically within reach.

4.3. Future directions

The present study did not include family-risk status, a commonly used predictor of reading difficulties, since the information obtained from families was incomplete and unreliable. While it is possible that classification accuracy could improve by including this information, such as in Puolakanaho et al.'s study (2007), other studies show that family-risk status is no longer relevant when preliteracy skills are included in the model (Thompson et al., 2015). Nevertheless, this should be tested in future work.

Even though Lexiland shows very high classification accuracy, its performance could potentially be improved by using other modelling techniques. Classification trees have shown great promise in improving predictive outcomes (Matsuki et al., 2016; Petscher & Koon, 2020). Moreover, classification trees are more amenable to non-experts such as parents and educators (interpretation of logistic regression coefficients is barely amenable to experts).

Additionally, any discretization of a continuous variable is to some extent arbitrary and, therefore, other cut-off thresholds for defining at-risk status should be explored. In the present work, we have labelled children 1.3 SD below the mean as *poor readers*. Scores in all reading tasks measured showed that these children effectively were not reading to the expected level by the end of first grade. However, a cut-off threshold of -1 or -1.5, which would increase or decrease the number of children in this group, could be explored. Model performance could change by using alternative thresholds, but specially the number of children labelled as in need of attention (or at least, in need for further assessment) would also change. The result of studying a set of cut-off thresholds could therefore inform public policy and remedial interventions.

To conclude, early identification of at-risk children is a feasible, inexpensive endeavour. Policy makers should think of it as an investment, the earlier the identification, the more successful the intervention. While early identification is paramount, it is not sufficient. If adequate remedial intervention is not set in place, then early screening loses its value. However, providing the educational system with quality easy-to-implement tools is a solid first step.

Chapter 4

Study Two: Cognitive substrates of reading acquisition

1. Introduction

Reading is a fundamental ability in modern societies, yet many children and adults struggle with reading. This has far-reaching negative consequences for personal development and professional achievement (Arnold et al., 2005). Thanks to decades of research, we now have a fairly comprehensive picture of the preliteracy skills required for successful reading. There is broad consensus regarding three critical skills: knowledge of letter sounds (letter knowledge; LK); rapid and efficient access to lexical representations (rapid automatized naming; RAN); and an ability to consciously manipulate the constituent units of oral language, generally referred to as phonological awareness (Boets et al., 2007; Lyytinen et al., 2006a; Muter et al., 2004; Schatschneider et al., 2004). Phonological awareness (PA) is a central construct in most, if not all, models of reading acquisition (for a meta-analysis see Melby-Lervåg et al., 2012). It is generally defined as a metalinguistic or metacognitive ability to identify and manipulate the sounds of a language. The term *sounds* in this context may refer to individual phonemes, syllables, or words. However, most research suggests phonological awareness at the phonemic level (*phonemic* awareness) is the most important component in reading acquisition (Castles & Coltheart, 2004). Typical PA tasks involve segmenting a word into its constituting syllables or phonemes, blending syllables or phonemes into a word, or replacing a syllable or phoneme within a given word or pseudoword. The important role of PA in reading is further confirmed by studies reporting that dyslexic children show a PA deficit (Melby-Lervåg et al., 2012; Vellutino et al., 2004), and that training in PA can improve reading skills (Bowyer-Crane et al., 2008). A fourth skill that taps into phonological processes—and also predicts early reading acquisition—but has received far less attention than PA, is verbal short-term memory. Verbal short-term memory (vSTM) has received a mixed treatment in the literature, sometimes treated as a foundational skill of interest (Moll et al., 2014; Ramus & Szenkovits, 2008; Torgesen et al., 1994) and sometimes as a covariate (Caravolas et al., 2012; Furnes & Samuelsson, 2010; Puolakanaho et al., 2007; Vaessen et al., 2010). Interestingly though, the core phonological deficits described in children with dyslexia often involve vSTM, alongside PA and RAN (Torgesen et al., 1994).

While all of the preliteracy skills mentioned above have been shown to play a role in predicting reading acquisition, PA is the most studied, both because of its central role, and because of its potential as a target for intervention. However, at least three aspects of the PA-reading relation remain unclear.

Universality. Most evidence comes from studies on English, whose orthography is atypical, in comparison with the orthographies of most other languages (Share, 2008). Orthographies can be characterized by the consistency of the mapping between graphemes and phonemes. In highly transparent orthographies, such as Spanish, Finnish, or Italian, the mapping between graphemes and phonemes is almost univocal, while in less transparent orthographies, such as English, this mapping depends heavily on the orthographic context in which the grapheme is embedded (Schmalz et al., 2015). Orthographic consistency, in turn, modulates the developmental trajectories of reading acquisition: high decoding levels are achieved faster in more transparent orthographies (Seymour et al., 2003). Moreover, several studies have shown that, PA together with RAN and LK skills account for larger amounts of variance in English than they do in other languages with more transparent orthographies (Caravolas et al., 2019; Moll et al., 2014). Therefore, it is still debated whether the central role attributed to PA in reading acquisition in English can be generalized to more transparent orthographies (Castles & Coltheart, 2004; Share, 2008; Verhoeven & Keuning, 2017). Questioning the central role of PA does not necessarily mean that PA has no role to play in predicting reading acquisition in languages with more transparent orthographies, but rather that its role might be less central with respect to other preliteracy skills (Duncan et al., 2013; Landerl et al., 2019; Verhoeven & Keuning, 2017). While each preliteracy skill adds unique variance that helps explain early reading skills, these skills are also correlated. For example, a child with strong PA skills may use this knowledge as a scaffold to learn letter names and vice versa: learning letter names can aid in the development of PA (Kim et al., 2010; Piasta & Wagner, 2010; Treiman & Kessler, 2004).

Causality. Reciprocal influences between PA and reading acquisition throughout development makes it harder to address the issue of the PA-reading relation (Castles & Coltheart, 2004; Charles Hulme et al., 2005). While children with reading difficulties often show accompanying poor PA skills, it is possible that these observed deficiencies in PA result from reduced or suboptimal reading experience. For example, Huettig and colleagues (2018) note the importance of distinguishing cause and effect when establishing the main factors that contribute to reading difficulties. They argue that in order to determine whether a given skill plays a causal role in reading development, it needs to be assessed in prereaders before any reading skills have developed. This rules out the possibility that the observed effects are a consequence of suboptimal reading experience rather than their primary cause. Longitudinal studies that initially test prereaders before reading instruction, although not conclusive, are thus a primary source of evidence to assess the causal role of PA skills in future reading performance (see also Goswami, 2015).

Operationalization. All the above aspects of the PA-reading relation have been additionally obscured by a third factor: the operationalization of PA (McBride-Chang, 1995; Runge & Watkins, 2006; Vloedgraven & Verhoeven, 2009; Yopp, 1988). Tasks used to test PA vary in difficulty on many dimensions. First, they may differ in terms of the linguistic unit of analysis, which could be whole-words, syllables, intra-syllabic units (onset and rimes), or phonemes. While it has been shown across orthographies that children's sensitivity develops along a trajectory from larger to smaller units (Anthony et al., 2011; Duncan et al., 2013; Papadopoulos et al., 2009; Ziegler & Goswami, 2005), it is still debated whether sensitivity to phonemes can be attained prior to any literacy exposure (Castles & Coltheart, 2004; Landerl et al., 2019). Longitudinal studies have used either measures of *phonological* awareness (including both syllabic and phonemic items) or *phonemic* awareness (only phonemic items) measures, further complicating the matter. Second, they may differ in terms of the kind of cognitive operation involved in the task. Both in English and Spanish, children are able to blend linguistic units before they can segment them, and identify them before they can manipulate them (e.g., detect identical onsets in two words vs. remove the onset) (Anthony et al., 2003). Different studies use a wide variety of different tasks, which makes it difficult to conduct comparisons across studies. Third, PA tasks vary in terms of the memory-load they impose. This has been identified as a crucial modulating factor in performance across phonological awareness tasks (Martinez Perez et al., 2012; Ramus & Szenkovits, 2008). Finally, PA tasks vary in their response format. PA is usually measured in tasks that require verbal responses, such as removing a given phoneme from a word and producing the resulting word/nonword. Naturally, producing a verbal response adds an additional cognitive process, which may or may not be tapping into the PA construct directly. For example, Cunningham and colleagues (2015) showed that grain size and response format in PA tasks constitute independent factors, and that the production of a verbal response contributed unique variance to decoding above and beyond the linguistic component involved. However, production of a verbal response is not usually defined as part of the PA construct, but as a means for measuring it. In sum, the wide range of task properties commonly used has presented a further obstacle in trying to reconcile divergences in the existing literature (Ramus & Szenkovits, 2008).

The above factors —universality, causality, and operationalization— may explain why evidence for the unique (i.e., above and beyond other variables) contribution of PA to reading has been inconsistent across orthographies. In the last two decades, several studies have attempted to address these issues by assessing preliteracy skills in prereaders in less transparent orthographies using longitudinal designs as well as cross-language approaches. While some of these studies have found evidence that supports a universal role for PA in reading acquisition (Caravolas et al., 2012; Furnes et al., 2019; Puolakanaho et al., 2008; Vaessen et al., 2010), others have not (De Jong & Van der Leij, 2003; Defior et al., 2008; Georgiou et al., 2012; Landerl et al., 2019; Mann & Wimmer, 2002; Schmitterer & Schroeder, 2019; Van Bergen et al., 2011)

There is an additional challenge in trying to make sense of this divergent evidence that has not, to the best of our knowledge, yet been systematically addressed: studies which show no evidence for a universal contribution of PA to reading acquisition also frequently report floor effects on PA measures (De Jong & Van der Leij, 2003; Georgiou et al., 2012; Landerl et al., 2019; Van Bergen et al., 2011). *Floor effects* are a form of scale attenuation encountered when measures are close to zero for most participants, thus providing an inaccurate measure of individual participant's ability. They generally result from tasks that are too difficult for the target participants, either due to poor item design or lack of adjustment for developmental stage. Most often, this is regarded as a methodological limitation in such studies. On the other hand, evidence in favour of a universal account has its own challenges. Some studies include a sample which already has some reading experience at study onset. As discussed above, this introduces a confound due to the reported reciprocal influences between phonological awareness and reading (Caravolas et al., 2012; Furnes & Samuelsson, 2010; Vaessen et al., 2010). In other cases, not all relevant covariates are included (crucially, verbal short-term memory and letter knowledge may be left out), making it difficult to compare the unique contribution of each predictor hard to perform (Furnes & Samuelsson, 2010; Puolakanaho et al., 2007). Finally, sample composition in these studies is often enriched with children at risk of reading failure, usually due to family history of dyslexia (Puolakanaho et al., 2008). Naturally, when trying to achieve a final sample that contains at least some children with reading difficulties, this is a sensible approach. However, it limits the generalizability of results to a broader, unselected population.

Understanding the unique contribution of PA to early reading skills across orthographies is relevant for both practical and theoretical reasons. An important practical implication is the design and use of appropriate screening tools for children at risk of reading difficulties. Early screening is vital, since it has been shown that remediation programmes are more effective the sooner they begin (Ozernov-Palchik & Gaab, 2016). If the unique contribution of PA is orthography dependant, then screening should be orthographic-specific (see for example, Solheim et al., 2020). Adaptation of tools developed for English-speaking children would not be appropriate for Spanish-speaking ones. From a theoretical standpoint, it raises new questions concerning the universality of the current prevailing model of reading acquisition. If the contribution of PA is not unique, does this mean PA has no role to play in reading acquisition? Can this explain the floor effects often reported in more transparent orthographies? If so, why are floor effects in PA often observed in more transparent orthographies but not in less transparent ones? We claim here that floor effects could be explained by a delayed development of PA skills in more transparent orthographies, rather than by measurement error. If, in more transparent orthographies, PA skills during the kindergarten years are only primitively or not at all developed, then it stands to reason that PA will show no unique contribution to later reading acquisition. Further, it is possible that other preliteracy skills will take its place. We believe LK is a strong candidate. Since, in more transparent orthographies, letter sounds are virtually equivalent to the phonemes they represent, in such orthographies LK might index children's ability to identify phonemes, thus replacing PA as a main contributor to later reading acquisition.

1.1. The present study

In the present study we examined the unique contribution of pre-reading phonological awareness to early reading skills in a transparent orthography, Spanish. Our hypothesis was that, in more transparent orthographies: i) delayed development of PA skills explain the previously observed floor effects of PA, ii) LK indexes children's ability to identify phonemes and thus takes a more central role in such orthographies.

To test this hypothesis, it was critical to design tasks that were sensitive to the general PA abilities of children at the time of testing. In order to tackle this issue, we employed a comprehensive assessment of phonological awareness, involving the manipulation of syllables and phonemes, that included four different tasks consisting of 163 items. We longitudinally assessed an unselected sample of children at two time points: in kindergarten, before any reading instruction has taken place, and at the end of Grade 1. Crucially, we computed latent ability scores through an item-response theory (IRT) approach, which allowed us to control for measurement error and compare tasks scores across different scales (Cole & Preacher, 2014; Hjetland et al., 2019). Moreover, in order to examine the unique contribution of PA relative to other preliteracy skills and general cognitive factors, we also assessed LK, RAN, and vSTM, as well as several other relevant control variables. At the end of grade 1, we repeated K5 measures and additionally measured reading skills. In order to account for the fact that children achieve high accuracy levels at the end of first grade (Seymour et al., 2003), we assessed decoding accuracy in words and pseudowords, as well as fluency, and modelled these factors independently.

2. Methods

2.1. Sample and measures

See Chapter 2 for a broad description of sample composition and measured cognitive variables. A brief overview is provided in the following section.

The sample was composed of children that completed Time 1 (K5) and Time 2 (G1) assessments (n = 388). Cognitive variables used in the present chapter include phonological awareness (PA), letter knowledge (LK), rapid automatized naming (RAN), verbal and non-verbal short-term memory (vSTM and nvSTM), vocabulary (VOC), intelligence quotient (IQ), and reading (decoding, fluency, and comprehension). Demographic variables included Age, Gender, and Socioeconomic status (SES). SES was defined as the highest achieved level of maternal education. It was treated as an ordinal variable with 3 levels: Low—Unfinished high school or less (n = 158)—, Middle—Completed high school (n = 64)—, and High—completed a Bachelor's degree or more (n = 91). Data was missing for 75 children (19.3 % of the sample), which were excluded from further analyses. All analyses were performed using R software (R Core Team, 2018).

2.2. Latent ability scores through item-response theory

Since the PA construct was measured by four tasks varying in terms of difficulty and cognitive load, we estimated a latent ability score for each child by combining all tasks measured in K5. This estimation of latent ability scores for PA using an item-response theory approach served two ends. First, it enabled us to directly compare difficulty levels among tasks. Second, it controls measurement error (Cole & Preacher, 2014; Hjetland et al., 2019). We estimated a 2PL model from the 163 phonological awareness items via the *ltm* package (Rizopoulos, 2006). Previous evidence shows that PA is a unitary construct, an assumption of IRT models (Anthony et al., 2011; Vloedgraven & Verhoeven, 2007). Model fit was assessed through comparison of the 2PL to a simpler Rasch model. Likelihood-ratio test confirmed that the additional discrimination parameter in the 2PL model significantly improved model fit to the data (LRT = 3370.87, df = 162, p < 0.001). Item fit was assessed through the *item.fit* function in the *ltm* package (Rizopoulos, 2006), which computes Yen's Q1 statistic . P values were obtained through 1000 Monte Carlo simulations. All items show excellent fit (all ps > 0.99). However, an examination of the difficulty parameters for each item showed that 12 items had extreme values. The difficulty parameter can be interpreted as the latent ability level where the expected proportion of correct responses is 0.5. Given that latent ability scores usually range from -4 to 4, values larger than 10 or smaller than -10 are very unlikely. Thus, items with difficulty values larger than absolute 10 were excluded. Excluded items belonged to the blending (1 phonemic item) and segmentation tasks (9 phonemic items, 2 syllabic items). A new reduced model was fit with the remaining 151 items. AIC, BIC, and log likelihood values all suggest the reduced model provides better fit than the complete one (AIC complete = 95155.22, AIC reduced = 88258.57, BIC complete = 96586.44, BIC reduced = 89584.43, LL complete = -47251.61, LL reduced = -43827.29). Using a likelihood ratio test to compare these two models was not appropriate, since they were fit using different data sets.

Person-level analysis. Having established adequate model fit, latent ability scores were computed for each child from the reduced 151 item model via Empirical Bayes through the *factor.scores* function in the *ltm* package (Rizopoulos, 2006). Pearson correlation coefficient between latent scores obtained from the complete model and from the reduced model was 0.99. Obtained latent ability scores were normally distributed around 0 (min: -2.9, max: 2.7). Finally, we assessed overall test information —an analogous to reliability in classical test theory— and confirmed that precision of measurement is centred around 0, suggesting that our PA tasks are most informative at average latent ability score levels.

3. Results

The rationale for the analysis was as follows. First, we studied the development of phonological awareness from K5 to G1, from raw scores in each task and time point, and from an IRT model to estimate difficulty and discrimination parameters, as well as latent ability scores. Next, we identified children who, during K5, could and could not read, and compared their preliteracy skills of K5 readers and non-readers. Finally, we tested our main hypothesis regarding the role of PA in reading acquisition in a transparent orthography using mixed effects linear models of decoding and fluency.

Table 4.1. Descriptive statistics for K5 measures and G1 reading

time		mean	SD	min	max	skewness	kurtosis	reliability	chance
K5	Age	5.82	0.29	6.34	7.36	0.03	-1.21	-	-
K5	IQ	10.09	5.42	0.00	28.00	0.60	0.08	0.88	-
K5	Vocabulary	0.83	0.12	0.27	1.00	-1.66	4.18	0.83	0.25
K5	non-verbal STM	3.57	1.23	1.00	6.00	-0.26	-0.53	0.63	-
K5	verbal STM	3.62	1.04	1.00	6.00	-0.05	-0.44	0.64	-
K5	blending phonemes	0.31	0.18	0.00	0.94	1.36	1.87	0.83^{+}	0.38
K5	blending syllables	0.83	0.15	0.21	1.00	-1.46	2.37	0.83^{+}	0.37
K5	onset matching phonemes	0.54	0.12	0.31	0.97	1.46	2.19	0.83^{+}	0.50
K5	onset matching syllables	0.59	0.15	0.19	1.00	0.81	-0.04	0.83^{+}	0.50
K5	rhyme pseudowords	0.54	0.16	0.00	1.00	0.34	0.99	0.83^{+}	0.50
K5	rhyme words	0.57	0.17	0.10	1.00	0.57	0.53	0.83^{+}	0.50
K5	segmentation phonemes	0.27	0.12	0.07	0.96	2.00	7.28	0.83^{+}	0.25
K5	segmentation syllables	0.41	0.17	0.09	1.00	1.36	2.13	0.83^{+}	0.33

K5	RAN colours	56.64	19.05	20.17	125.10	1.22	1.91	-	-
K5	RAN objects	48.53	12.60	24.92	97.72	0.90	1.30	-	-
K5	letter name	0.60	0.23	0.09	1.00	0.14	-1.18	0.91^{+}	0.33
K5	letter sound	0.55	0.21	0.14	1.00	0.32	-0.85	0.91^{+}	0.33
G1	decoding words	0.75	0.34	0.00	1.00	-1.32	0.22	-	-
G1	decoding pseudowords	0.68	0.32	0.00	1.00	-1.14	-0.09	-	-
G1	fluency	21.13	15.98	0.00	99.00	1.13	2.38	-	-

Units: Vocabulary, Blending, Onset matching, Rhyme, Segmentation, Letter and Decoding: mean accuracy; IQ, non-verbal STM, verbal-STM: maximum level achieved; RAN: total response time; Fluency: words read correctly per minute. Reliability is Cronbach's alpha. $^{+}$Reported reliability corresponds to the composite score.

Descriptive statistics for K5 measures and G1 reading are reported in Table 4.1 (See Appendix B for other G1 measures). Chance denotes the chance level for each task that involved a multiple-choice response format. Composite measures were computed for the two RAN tasks, for the two LK tasks, and for the two decoding tasks (RAN r = 0.56, CI 95% 0.49 – 0.63, p < 0.001; LK r = 0.77, CI 95% 0.73 – 0.81, p < 0.001; decoding r = 0.96, CI 95% 0.95 – 0.96, p < 0.001). Correlations among all variables measured in K5 and reading measured in G1 were studied to assess collinearity issues for model building (Table 4.2). The strongest correlations among K5 measures were between LK and PA, LK and vocabulary, and LK and verbal short-term memory. The strongest correlations between K5 variables and G1 reading were for LK, followed by RAN and non-verbal STM. All correlations were significant at the 99% level with p values corrected to through false-discovery rate.

Table 4.2. Pearson correlation coefficients for K5 variables and G1 reading measures

	time	var	1	2	3	4	5	6	7	8
1	G1	decoding								
2	G1	fluency	0.67****							
3	K5	PA	0.26****	0.19***						
4	K5	IQ	0.21****	0.19***	0.28****					
5	K5	Voc	0.27****	0.16**	0 .27****	0.27****				
6	K5	nvSTM	0.36****	0.28****	0.26****	0.26****	0.27****			
7	K5	vSTM	0.38****	0.32****	0.27****	0.26****	0.24****	0.35****		
8	K5	RAN	-0.38****	-0.34****	-0.15**	-0.24****	-0.28****	-0.31****	-0.28****	
9	K5	LK	0.50****	0.50****	0.36****	0.31****	0.38****	0.31****	0.44****	-0.34****

**** p < 0.0001,*** p < 0.001, ** p < 0.01 false discovery rate correction

PA: phonological awareness, Voc: vocabulary, nvSTM: non-verbal short-term memory, vSTM: verbal short-term memory, RAN: rapid automatized naming, LK: letter knowledge

3.1. Development of phonological awareness

3.1.1. Raw scores

In order to evaluate performance on each task, we first performed one-sample t-tests of raw accuracy scores against chance, since all tasks were presented in a multiple-choice format (Figure 4.1). Children performed better than chance across all tasks ($p < 0.001$), except for blending phonemes, where average performance was significantly below the chance level (mean = 0.31, chance = 0.37, 95% CI = (0.29, 0.33), t = -7.0387, df = 387, $p < 0.001$). Notably though, performance in the other two PA tasks involving phonemes was barely above chance (segmentation phonemes = 0.27, chance = 0.25, 95% CI = (0.26, 0.28); onset matching phonemes = 0.54, chance = 0.5, 95% CI = (0.53, 0.55)).

Next, since PA skills were assessed both in K5 and G1, we could evaluate growth in PA skills across time. A linear mixed effect model with raw accuracy as the outcome and task, time, and task-time interaction as predictors showed significant effects for all predictor variables, including the time-task interaction. Post-hoc comparisons for each task across time points showed significant improvements in accuracy for all tasks (all $p < 0.001$, corrected through false discovery rate).

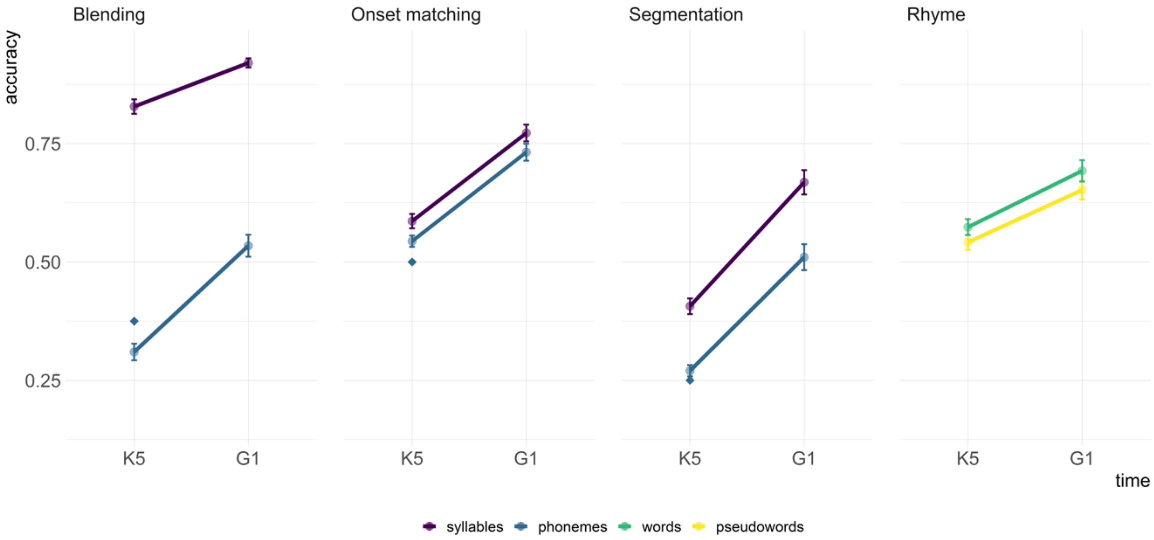


Figure 4.1. For all tasks, syllabic performance was significantly better than phonemic performance. Phonemic performance was barely above chance levels for Onset matching and Segmentation and significantly below chance level for Blending. All PA skills improve with time from K5 to G1. Error bars represent 95% confidence interval. Diamonds represent chance levels for tasks involving phonemes.

3.1.2. Latent ability scores

Task *difficulty* was examined for each task and grain size (that is, syllabic vs. phonemic items). Overall average difficulty was 0.5. Tasks arranged from less to more difficult were: blending (mean = -1.23) < rhyme (mean = 0.33) < onset matching (mean = 0.61) < segmentation (mean = 1.87). Pairwise comparisons through two-sample t-test (with p values corrected through Tukey method) showed significant differences between blending and onset matching (t(147) = 1.84, p < 0.05) and between blending and segmentation (t(147) = 3.1, p < 0.001). With respect to grain size, syllabic items were significantly less difficult than phonemic ones (mean syllables = -0.62, mean phonemes = 1.65, t(129) = 2.27, p < 0.001). These results are consistent with the expected progression of development of phonological awareness from syllabic to phonemic units, and from blending to identifying to segmenting (Anthony et al., 2003, 2011; Ziegler & Goswami, 2005).

Regarding *discrimination* parameters, average discrimination was 0.3, with tasks arranged from less to more discriminative: segmentation (mean = 0.24) < blending (mean = 0.35) < rhyme (mean = 0.38) < onset matching (mean = 1.16). Pairwise comparisons showed significant differences between segmentation and onset matching (t(147) = 0.92, p < 0.001), blending and onset matching (t(147) = 0.80, p < 0.001), and rhyme and onset matching (t(147) = 0.77, p < 0.001). No significant differences were observed in discrimination parameters between syllabic and phonemic items.

Taken together, the results show better performance at the syllable than at the phoneme level, which scores at chance or barely above chance for the phoneme level, and an overall growth in performance from K5 to G1. Moreover, PA tasks showed adequate difficulty and discrimination parameters.

3.2. Preliteracy skills and reading status

In order to assess the unique contribution of PA to reading *before any reading experience*, children were tested on their reading levels in K5 through a list of 15 words and 15 pseudowords. Children are not expected to have reading skills at this stage as reading is not explicitly taught in kindergarten. Accordingly, 86.3% of the sample could not decode any pseudowords, while only 11.3% correctly decoded more than 10 pseudowords. In order to use a conservative criterion, we defined *K5 readers* as those that decoded one or more pseudowords correctly, which constituted 13.9% of the sample. We used pseudoword decoding as a criterion for reading; it is more conservative measure than word reading because it excludes the use of any familiar whole-word recognition strategies.

Following the vast literature on the role of preliteracy skills in reading acquisition, we compared K5 readers vs. non-readers in each preliteracy skill using one linear regression model per task, with task score (for PA and LK) or response time (for RAN) as outcomes, and Age, IQ and group (K5 reader vs. non-reader) as predictors (Figure 4.2). In all models, the group coefficient was significant at the 99% confidence level. Planned comparisons of marginal means showed that *K5 readers* outperformed non-readers in all preliteracy skills. All K5 readers were removed from further analysis in order to avoid reciprocal effects of PA and reading.

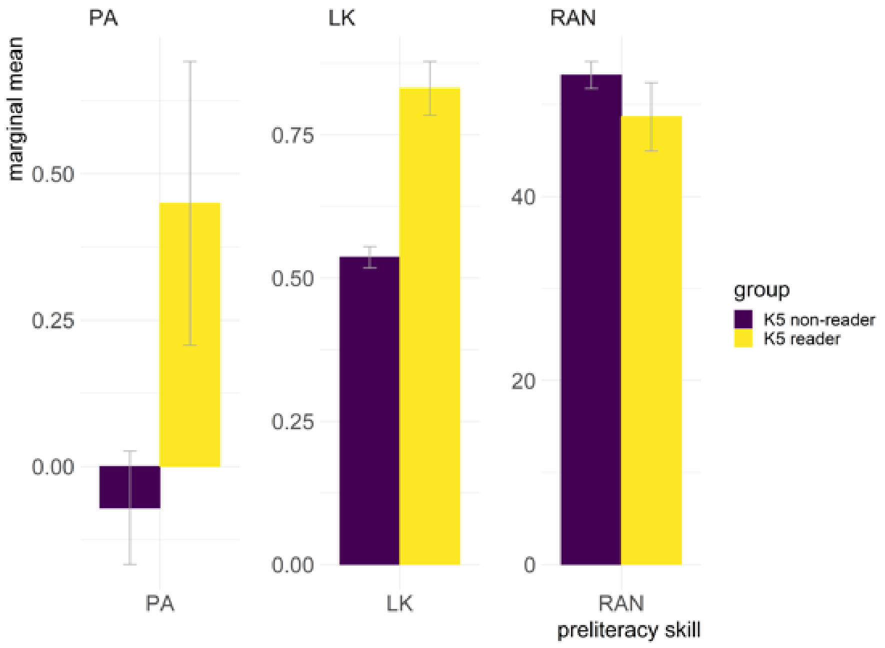


Figure 4.2. Preliteracy skills performance of K5 readers (n = 54) vs non-readers (n = 334). Marginal means, controlling for Age and IQ. Error bars represent 95% confidence interval. Marginal means represent latent ability scores for PA, mean accuracy for LK, and response times in seconds for RAN (smaller scores mean better performance). K5 readers outperform K5 non-readers across all measures.

3.3. Unique contribution of PPA to future reading skills

We evaluated the unique contribution of PA to early reading abilities, while controlling for relevant covariates, by running linear mixed effects regression models with preliteracy skills measured in K5 as predictors (LK, RAN and PA), and two outcome variables: decoding (composite of words and pseudowords accuracy) and fluency (words read per minute) measured in G1. For PA, we used latent ability scores from an IRT model including syllabic and phonemic items (see section 2.2 in the present chapter). While phonemic awareness is generally reported to be the main predictor of early reading skills —as opposed to syllabic awareness— and is used in many studies (Caravolas et al., 2012; De Jong & Van der Leij, 2003; Holopainen et al., 2014; Van Bergen et al., 2011; Ziegler et al., 2010), the literature remains unclear in this respect, since some studies use a combination of both, either explicitly or implicitly (Furnes & Samuelsson, 2010; Georgiou et al., 2012; Landerl et al., 2019; Puolakanaho et al., 2007). Given the low performance of children on the phonemic items, we decided to use a combination of phonemic and syllabic items. All reported models were also fit with latent ability scores for phonemic awareness and syllabic awareness separately. Results for the full model remained the same and are thus not reported. School was included as random intercept to account for the nesting of children across schools. Age, Gender, IQ, Vocabulary, vSTM, nvSTM, and Maternal Education, as a proxy for socioeconomic status (SES), were included as control variables. vSTM was treated as a control variable in order to focus on the core component of the PA construct and because of the large memory load involved in some of the PA tasks. Since PA and reading have shown reciprocal effects (Castles & Coltheart, 2004), all children that showed any reading skill in K5 were excluded from the analysis. For this reason, we refer to PA skills in these children as PPA (pre-reading phonological awareness). For model specification and selection, we followed Meteyard and Davies (2020) recommended practices on linear mixed-effects models. First, a *null model* containing only a random intercept for School was fitted. No random slopes were added since the number of children by school was low for estimation purposes. Model building continued from minimal to maximal. In the next step we computed the *preliteracy model*, adding three preliteracy skills of interest as fixed effects: PPA, LK, and RAN. Finally, we ran the *full model*, in order to assess the unique contribution of preliteracy skills *after controlling for relevant covariates,* adding all covariates as fixed effects (Age, Gender, SES, IQ, Vocabulary, vSTM and nvSTM). Model details are available in in Appendix B.

The *null models*, containing only the random effect for School explained approximately 10% of the variance in decoding and 6% in fluency. In the *preliteracy models*, LK, RAN, and PPA all contributed uniquely to decoding. LK and RAN, but not PPA, contributed uniquely to fluency. All variables combined explained 39% of the variance in decoding, and 31% of the variance in fluency. Both models (accuracy and fluency) significantly improved model fit as compared to the null model. In the *full models* (Figure 4.3), which included all relevant covariates in addition to preliteracy skills, LK and RAN still contributed unique variance among preliteracy skills (see Table B.2 in Appendix B for further details). Crucially, PPA no longer contributed unique variance to decoding. In other words, once covariates were included, the unique contribution of PPA was no longer significant. Among covariates, vSTM, nvSTM, and SES all contributed unique variance to decoding. For fluency, nvSTM and Gender were unique predictors (with boys outperforming girls). Overall, the full models accounted for 45% of the variance in decoding and 38% of the variance in fluency. As for variance explained by each predictor of interest while keeping all other variables constant, for decoding, PPA contributed 2.2% of additional unique variance, LK contributed 6.4%, and RAN 6.0%. For fluency, PPA contributed 0.3%, LK 7.7%, and RAN 4.8%. Both full models (accuracy and fluency) significantly improved model fit as compared to the preliteracy skills models (see Table B.3 in Appendix B for further details).

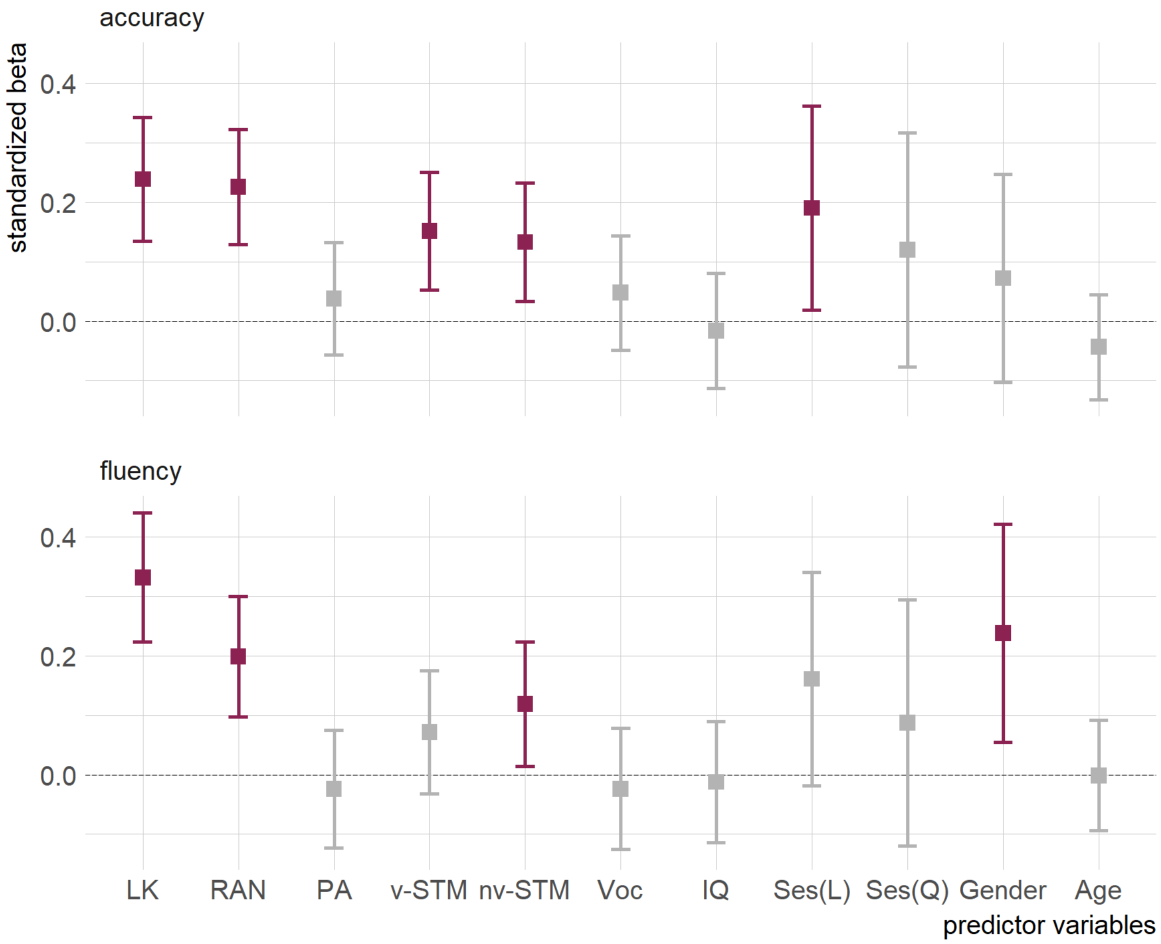


Figure 4.3. Regression coefficients for the full prediction model of reading from preliteracy skills while controlling for relevant covariates. Prediction model coefficients for decoding (top panel) and fluency (bottom panel). School was included as a random intercept (not shown). Error bars represent 95% confidence intervals. Colour shows significant predictors for each model (different from zero). RAN coefficients are reversed for illustration purposes. For SES, since it is an ordinal variable, L indicates a coefficient for a linear term, and Q for a quadratic term.

For the decoding model, the lack of a PPA effect in the presence of covariates was further examined. We reasoned that if the effect of PPA on reading was modulated by any of these control factors, as evidenced by the change in the model coefficient for PPA, interaction effects were likely. Thus, we estimated three new models including interaction terms between PPA and verbal short-term memory (model 1), PPA and non-verbal short-term memory (model 2), and PPA and SES (model 3). The only significant interaction effect observed was for PPA and SES (see Table B.4 in Appendix B for further details). An examination of the pattern of reading-PPA relations by SES group showed that the PPA-reading relation was stronger for the low than the middle and high SES groups. This new model significantly improved model fit over the full model without any interaction terms (delta r squared = 3.1%, LRT Chi Sq (2) = 13.69, $p < 0.001$).

4. Discussion

In the current study we assessed the unique contribution of PA to early reading skills in a transparent orthography. By computing latent ability scores from a comprehensive PA battery, we overcame the floor effects of PA often reported for more transparent orthographies. In two regression models of decoding and fluency, we showed that pre-reading phonological awareness (PPA) does not uniquely contribute to early reading acquisition above and beyond other preliteracy skills, while controlling for several relevant covariates. Instead, we showed that LK and RAN (and vSTM in the case of decoding) are the most relevant predictors of early reading skills. Importantly, our prediction models accounted for large amounts of variance (38% and 45%) even in the absence of a significant unique contribution from PPA. Our findings shed light on how the dynamic interplay among preliteracy skills may reveal itself across orthographies.

4.1. Development of PA in a transparent orthography

As reported in studies of PPA in prereaders in more transparent orthographies, phonemic awareness showed floor effects (at chance or barely above chance levels) in our sample, as evidenced by average scores and by difficulty parameters in the item-response theory model. Floor effects have been a main explanatory reason for not finding a unique contribution of phonemic awareness to reading in more transparent orthographies (De Jong & Van der Leij, 2003; Georgiou et al., 2012; Landerl et al., 2019; Van Bergen et al., 2011). However, while this argument makes methodological sense –one would expect no significant unique contribution when the predictor does not show sufficient variance— its theoretical interpretation should not be dismissed. Why is it common to see floor effects in phonemic awareness measures in kindergarten children from languages with more transparent orthographies? In line with previous studies, our results suggest that that *phonemic awareness* develops "late but fast" in transparent orthographies (Defior et al., 2008; Mann & Wimmer, 2002).

4.2. Unique contribution of PPA to reading acquisition

Results from the full regression models for both decoding and fluency show that PPA does not contribute uniquely to reading acquisition above and beyond other preliteracy skills when critical covariates are included. The comprehensive assessment and large sample size in our study confirm that the null unique contribution from PPA was not a result of measurement error or lack of power. These findings add converging evidence from a Spanish speaking population to the available studies on more transparent orthographies such as Dutch, German, Finnish and Greek (De Jong & Van der Leij, 2003; Defior et al., 2008; Georgiou et al., 2012;

Landerl et al., 2019; Mann & Wimmer, 2002; Schmitterer & Schroeder, 2019; Van Bergen et al., 2011). On the other hand, these results contradict those reported by Caravolas and colleagues (2012) in their longitudinal crosslinguistic study including Spanish. A possible explanation for the discrepancy is that in their study children had some reading experience at study onset. This could have prompted the development of PA. The present results, in contrast, come from a sample of children who, at study onset, could not decode any pseudowords; therefore, no reciprocal effects were expected. The reciprocal effects of reading on the development of PA could unfortunately not be tested in the present sample since the proportion of readers at study onset was very low (13%). This did not warrant inclusion of an interaction term in the model, nor building a separate model specifically for those children. Additionally, in our study, unlike that by Caravolas et al., we report a significant unique contribution from pre-reading vSTM to reading. In their study, Caravolas et al. (2012) cite the low reliability of vSTM as an explanatory factor, noting it did not make a unique contribution to reading. This suggests they may have found a pattern of results similar to ours if the vSTM measure had been more reliable in their study. Also, the decoding measures used in their study and ours differed considerably. With regard to other more transparent orthographies, results on Finnish are also pertinent for our findings; since like Spanish, Finnish can be categorized at the extreme of orthographic consistency. In a study reported in Puolakanaho et al. (2007), preliteracy skills were compared in a sample of 200 children from 3.5 years of age, half of whom had a family history of dyslexia. Although they reported PA as a longitudinal predictor of reading skills in pre-reading children, this effect was only observed at a time point where RAN was not measured. At the other two time points, in which RAN was measured, PA did not show any effect above LK and RAN. Moreover, differences in sample composition between their study and ours likely had consequences for the findings. The Finnish sample was enriched by children with a family risk of dyslexia, while the present study was composed of an unselected sample of children.

The sum of evidence from longitudinal studies on more transparent orthographies thus casts doubts on a universal role for PPA during reading acquisition. Having established that PPA does not contribute unique variance to explaining early reading acquisition, we should ask if PPA has any role to play in such reading acquisition. Landerl and colleagues (2019) have put forward an account based on their results from a crosslinguistic longitudinal study of preliteracy skills in English, French, German, Dutch, and Greek. Having found a complex pattern of prediction across orthographies, they propose that PA in more transparent orthographies may develop as a corequisite rather than as a prerequisite of reading acquisition. We believe that a different, tighter association between PPA and LK can accommodate the observed pattern. Following Mann & Wimmer's (2002) thesis, in line with the proto-literacy hypothesis (Barron, 1991), “phoneme awareness must be triggered by something above and beyond the experiences that are sufficient to support primary language development” (2002, p. 676). That “something” might come from explicit letter name/sound instruction or from explicit phonological awareness activities. In the former case, at an initial point in time, we should see LK as a main predictor of future decoding and none or only a small unique contribution from PPA. In the latter, we would see a main role for PPA. From an interactive LK-PA standpoint (Charles Hulme et al., 2005; Kim et al., 2010; Piasta & Wagner, 2010) both skills should develop later on. This account would seem to suggest that the differences observed in prediction patterns for decoding are just a matter of differences in kindergarten instruction or home literacy environments across countries. However, a further point can be made. When both skills are present, their relative contribution differs across orthographies based on the amount of information they convey (Vousden et al., 2011). In less transparent orthographies, where the number of phonemes tends to be larger than the number of graphemes used to represent them, the ability to identify and manipulate phonemes (i.e., PA) has larger explanatory value than knowing the letters. Additionally, in such orthographies, knowledge of letter sounds is not enough to correctly sound out words. Therefore, in a predictive model, both skills will contribute significant and independent amounts of variance to explaining early reading acquisition. On the contrary, in more transparent orthographies, given the almost one to one mapping between graphemes and phonemes, letter sounds are virtually equivalent to the phonemes they represent. As pointed out, in reference to Finnish, “*Because the Finnish language is so transparent, letter sound knowledge and phonemic awareness are near synonymous, and consequently, once mastery of the alphabetic principle, i.e., sounds of the letters, has been achieved, reading is underway*” (Lyytinen et al., 2015, p. 334). In this case, LK indexes children’s ability to identify phonemes, thus replacing PA as a main contributor to later reading acquisition.

In sum, the unique contributions of PA and LK as longitudinal predictors of decoding abilities is the result of a combination of kindergarten instructional practices and the home literacy environment, as well as the differential information content contributed by LK and PA across orthographies.

4.3. PA tasks: response format and procedure

An additional difference between this and previous studies is the operationalization of PA. Probably, the most critical difference stems from response formats. The PA construct is frequently measured through verbal responses, while in our tasks all responses were given in a multiple-choice format. Two points need to be considered when analysing this difference. First, despite the change in response format, we successfully replicated the developmental trajectories and the difficulty pattern reported in previous studies (development from larger to smaller units; blending easier than segmenting), both within and across testing times. Second, as stated before, Cunningham and colleagues (2015) have shown that producing a verbal response explains unique variance in the PA-reading relation, above and beyond that explained by comparison measures (the same task) with no verbal response. Clearly, this additional dimension of PA is lacking in our study. However, we see no reason, in principle, to include a verbal response as part of the core construct of PA. Also, by displaying response options on screen (and accompanying auditory stimuli with a visual representation) we have substantially decreased the memory load involved in solving the task. Thus, we have strong grounds to claim that our PA tasks are tapping into the PA construct, albeit through a different measurement.

4.4. PA and verbal short-term memory

A surprising finding from this study was the relevant role that pre-reading vSTM plays in the prediction of decoding skills. We originally included vSTM as a covariate, in order to control for the large memory load that PA tasks place on participants. However, as stated earlier, vSTM belongs to the broader construct of phonological skills important for reading acquisition, which includes PA and RAN in addition to vSTM. Hence, vSTM it is sometimes treated as a preliteracy skill *per se* (Moll et al., 2014; Ramus & Szenkovits, 2008; Torgesen et al., 1994), sometimes treated as a covariate (Caravolas et al., 2012; Furnes & Samuelsson, 2010; Puolakanaho et al., 2007; Vaessen et al., 2010), and sometimes treated as a single phonological construct together with PA (Knoop-van Campen et al., 2018; Martinez Perez et al., 2012; Moll et al., 2014). The present results suggest that vSTM predicts reading skills above and beyond other preliteracy skills and other general cognitive factors. We argue that this result can be explained by the underlying cognitive operations involved in learning to read in a transparent orthography. As stated before, given the almost one to one mapping between graphemes and phonemes, and thus the strong information content of letter sounds, converting each grapheme into its corresponding phoneme is almost trivial when there is advanced knowledge of letter sounds. Once this first step has been achieved, the next most critical operation is maintaining these letter sounds in memory to blend them. Thus, in more transparent orthographies, strong letter knowledge and memory skills are paramount for successfully acquiring early reading skills.

4.5. Limitations

Languages vary not only in their orthographic consistency but also in properties of the oral language itself, such as the rhythm of their syllabic structure. It is possible that these, less explored, properties also influence the development of PA and thus the PA-reading relation. For example, rhythmic properties vary in stressed-timed languages and syllable-timed languages, such as English and Spanish respectively. Rhythm, in turn, has recently been given more attention in defining the process of speech segmentation, which, in turn, affects the development of phonological skills (Wood & Connelly, 2009). While these linguistic properties have been much less explored, a provocative study across six alphabetic orthographies varying in consistency, syllabic structure, and rhythm found rhythm explained differences in the development of phonological awareness better than orthographic consistency (Duncan et al., 2013). The role of these other linguistic properties should be further explored in order to better understand how they interact with orthographic consistency to modulate the development of PA and reading acquisition.

A second limitation is the lack of information on teaching practices. While assessment of teaching practices was beyond the scope of the present study, there is large variability in the methods used for teaching reading in Uruguay. We are aware that variations in teaching methods might impact both the development of PA skills and the PA-reading relation. Including teaching practice as an additional variable in our model might shed further light on the conditions under which PPA uniquely contributes to reading acquisition and how this is modulated by teaching practices.

Finally, an additional factor that needs to be considered is the fact that children were tested in groups, which could lead to less focused attention and, consequently, impaired understanding of the instructions. However, the reliability of the tasks, the correlation matrix, and the developmental trajectories observed, suggest that children did understand the instructions and tried to complete each task to the best of their capacity.

To summarise, we found PPA made no unique contribution to later reading acquisition in a transparent orthography. These results cannot be explained by measurement error in PA, as has been cautioned with respect to previous studies. Instead, we found that the strongest contributors to decoding were RAN, LK (and vSTM), while the strongest contributors to fluency were RAN and LK. We propose that a delayed developmental trajectory for PA, a strong role for vSTM, and a tighter association between LK and PA —influenced by home literacy and educational practices as well as the intrinsic characteristics of the orthographic system— can accommodate these and previous results.

Chapter 5

Study Three: Dissecting the contribution of rhythmic sensitivity

1. Introduction

1.1. Rhythm and speech

Phonological awareness is a core skill for the acquisition of reading. But, what is the cognitive basis for the development of phonological awareness? The very concept of phonological awareness arises in the context of the studies by Isabelle Liberman and colleagues in the seventies, in trying to understand the basic unit of speech segmentation (Liberman et al., 1974): "If a writer is to represent a segment of whatever kind or size, he must first have succeeded in explicitly abstracting it from the acoustic stream of speech" (1974, p. 202). Speech segmentation has thus been considered central to the understanding of the development of phonological representations[3].

Early foundational work by Anne Cutler and Jacques Mehler proposed that speech segmentation is based on the analysis of the rhythmic structure of the speech input (Cutler et al., 1986; Cutler & Mehler, 1993; Mehler et al., 1981). The general idea is that a listeners' attention is biased towards stress in speech—giving rise to the perception of rhythm or prosody—which aids in the segmentation of the continuous speech stream. In English, a stress-timed language, the perception of rhythm arises from stressed syllables which are equally spaced in time, with a varying number of unstressed syllables in-between. English speakers use syllable stress as a cue for identifying word boundaries as words in English most commonly start with a stressed syllable. In Spanish, a syllable-timed language, listeners perceive all syllables to be of equal duration, and the perception of rhythm arises from these equally spaced syllabic units. Thus, in Spanish, syllables themselves are used to segment speech (Cutler et al., 1992). In any case, it is *rhythmic sensitivity*—that is, sensitivity to the rhythmic patterns present in the speech signal—that underlies speech segmentation and, thus, the development of phonological representations. Within speech stimuli, rhythmic sensitivity generally refers to the perception of changes in stress in an acoustic signal—i.e., prosody. Prosody is a phonological subsystem of speech that entails stress, timing, and intonation of segmental (phonemes) and suprasegmental (syllables, phrases, utterances) units of speech. The

---

[3] With the term phonological representations, we mean the abstract representation of speech sounds that arise as the product of speech segmentation. These include information of phonemes, syllables and stress in words, which interface speech processing during perception and production. Phonological awareness, in particular, refers to the conscious access and manipulation of phonological representations. We will use the term phonological processing as an umbrella term to refer to phonological awareness and verbal short-term memory combined.

combination of stress and timing gives rise to the perception of rhythm, which in the acoustic signal is embodied as the amplitude envelope. Therefore, rhythmic sensitivity in linguistic stimuli generally refers to sensitivity to stress or stress' fluctuations in time—i.e., the amplitude envelope.

Contemporary accounts of auditory processing provide mechanistic explanations of the rhythm-speech connection. At the neural level, this relation can be understood through the role that endogenous cortical oscillations play in speech segmentation. In the brain, oscillations in the auditory cortex show increased power in response to speech in the delta (1 to 3 Hz), theta (4 to 8 Hz), and gamma bands (30 to 60 Hz) (Baroni et al., 2019; Ghitza & Greenberg, 2009; Giraud & Poeppel, 2012; Luo & Poeppel, 2007). These frequency bands roughly correspond to prosodic, syllabic, and phonemic rates in speech. This "coincidence", which is likely the result of a common mechanism for speech production and perception (Morillon et al., 2010), is exploited by the brain to parse the speech input. Synchronization of endogenous cortical oscillations with the speech envelope in the delta and theta rates allows the brain to segment the continuous stream of information into discrete units. This mechanism could explain not only how syllabic information is extracted, but also how phonemic information is obtained. The rationale is that slower and faster oscillations are nested, with slower rhythms' phase modulating faster rhythms' amplitudes. Therefore, synchronization at lower frequencies would also affect processing at higher ones in the gamma range, which broadly correspond to phonemic rate (Giraud & Poeppel, 2012).

An influential study carried out by Nina Kraus and colleagues (Woodruff Carr et al., 2014) directly linked neural synchronization, phonological processing, and rhythmic sensitivity in a sample of prereaders (see Colling et al., 2017 &; A. Tierney & Kraus, 2013, for similar paradigms with older children and adolescentes). By assessing neural synchronization and rhythmic sensitivity in the same children, the study tested the hypothesis that rhythmic sensitivity measured behaviourally is directly linked to the neural encoding of speech rhythm. In order to assess rhythmic sensitivity, children were asked to synchronously tap to a beat produced by a researcher at two frequencies (1.67 and 2.5 Hz). For neural encoding measures, children listened to *ba*, *da*, and *ga* syllables, while their brain activity was recorded using EEG. Additionally, children were assessed on prereading skills (phonological awareness, RAN, and auditory short-term memory) and musical perception. Children were classified into a synchronizer and non-synchronizer group according to performance in the tapping task. The groups showed differences in phonological awareness, auditory short-term memory, rapid automatized naming, and musical perception. Within the synchronizer group (n = 22), non-linguistic rhythmic abilities correlated with precision in the neural encoding of speech stimuli. In a more ambitious analysis, a hierarchical regression model showed that rhythmic perception and production significantly predicted variance in neural encoding above and beyond prereading skills and general cognitive measures in the synchronizer group. These findings led to the conclusions that: i) non-linguistic rhythmic abilities can be used as an indirect measure of precision of neural synchronization to auditory stimuli, and ii) rhythmic sensitivity contributes to the development of phonological processing in prereaders.

1.2. Rhythm and reading

If rhythmic sensitivity impacts phonological processing, and phonological processing is key for reading acquisition, then it comes to reason that rhythmic sensitivity should contribute to early reading skills. The Temporal Sampling Framework proposes that the precision in brain synchronization to speech at different rates can explain the development of reading skills through the mediation of phonological awareness (Goswami, 2011). In a similar vein, Wood and colleagues extended Mehler and Cutler's work proposing that rhythmic sensitivity (prosodic awareness in their terms) contributes to reading skill more generally through phonological and morphological awareness (Wood & Connelly, 2009).

Much evidence has been accrued on the contribution of rhythmic sensitivity to reading. A common paradigm in these studies consists of presenting rhythmic stimuli to participants and linking their performance to reading abilities. Stimuli commonly consist of low pass-filtered sentences that retain only suprasegmental phonological information (i.e., prosody) but no segmental information (see, for example, Holliman et al., 2017). Since working with naturalistic stimuli can often interfere with the isolation and manipulation of specific components of the signal, a parallel approach in studying rhythmic sensitivity has used artificial non-linguistic stimuli. Non-linguistic rhythmic tasks generally involve asking participants to repeat a sequence of taps or tap along with a beat, where inter-tap timing or sequence length are variable (see, for example, Tierney & Kraus, 2013). Both paradigms have found strong associations between rhythmic sensitivity and reading abilities, both in stressed-timed languages (Anvari et al., 2002; Goodman et al., 2010; Holliman et al., 2017; Ozernov-Palchik et al., 2018; Steinbrink et al., 2019) and in syllable-timed languages in a smaller number of studies (Calet et al., 2015; Lundetræ & Thomson, 2018; Protopapas et al., 2006). The rhythmic characteristics of the studied language are particularly relevant in this framework, given the differential role that stress plays in the perception of rhythm in speech, as explained earlier.

However, most evidence comes from older children past the decoding stage, which limits the findings in at least two ways. First, at later stages of reading acquisition, during the development of reading fluency and comprehension, rhythmic sensitivity could play a role at longer time scales corresponding to prosodic reading, but not through phonological processing at shorter time scales (Kuhn et al., 2010). According to the presented frameworks, the rhythm-reading link should be observed at early stages of reading acquisition, where the role of segmental phonological processing is strongest (Muter et al., 2004). Additionally, given the known reciprocal effects of reading on phonological representations (Castles & Coltheart, 2004), the rhythm-phonology connection should thus be assessed prior to any reading experience. Second, formal musical experience modulates rhythmic sensitivity. Thus, the earlier it is assessed, the more likely it is that rhythmic sensitivity reflects basic phonological processing rather than learned abilities influenced by experience. Therefore, it is particularly important to address the rhythm-phonology-reading link using evidence derived from longitudinal studies involving prereaders. Recently, Lundetræ and Thomson (2018) followed 479 Norwegian children from school entry to the end of first grade to assess the contribution of a rhythmic task in classifying spelling and reading status above and beyond regular prereading skill assessments. Their rhythmic task consisted of drumming synchronously to a beat on a tablet at 1.5 and 2 Hz. Results showed that rhythmic abilities at 1.5 Hz improved classification accuracy for spelling and only marginally for reading (they did not analyse the contribution of non-linguistic rhythm at 2 Hz). Apart from providing further support for the role of rhythm in reading and spelling, this study provided novel evidence for the role of non-linguistic rhythm in a transparent orthography. However, Spanish and Norwegian differ in their prosodic features, with Spanish being a syllable-timed language and Norwegian a stressed-time language like English.

To the best of our knowledge, there is only one longitudinal study assessing the connection between rhythmic sensitivity and early reading in a syllable-timed language (Calet et al., 2015). Spanish speaking children were followed longitudinally from kindergarten to second grade. They were assessed on linguistic stress perception (at words and sentences) and non-linguistic rhythm production through a reproduction task. Their results show that non-linguistic rhythm repetition abilities during kindergarten predict reading acquisition one year later, above and beyond phonological awareness, IQ, and vocabulary. Lexical and metrical stress tasks during kindergarten did not show any link with future reading skills. The authors argue that the small sample size and the low reliability in lexical and metrical stress tasks could explain the lack of significant effects.

1.3. The present study

In the present study, we further characterized the contribution of rhythm to reading acquisition. With this aim in mind, we tested rhythmic sensitivity at three different frequencies and phonological processing before reading onset, relating them to reading efficiency at the end of first grade. We contribute to the current gaps in the literature by running a longitudinal study starting with prereders in a sample of Spanish (a syllable-timed language) speaking children.

2. Methods

See Chapter 2 for a broad description of sample composition and measured cognitive variables. A brief overview is provided in the following section.

2.1. Participants

The sample was composed of children attending K5 in public schools in Montevideo. A total of 442 children completed the rhythmic sensitivity task and are thus included in the present study. Children were tested at the school, in groups of four to five.

2.2. Demographic and cognitive measures

Cognitive variables included in the present analysis include demographic (Age, Gender and School), general cognitive (IQ and non-verbal STM), and phonological processing (verbal STM, phonological awareness at the syllable (PAS) and phoneme level (PAP)) measures.

2.3. Reading (efficiency)

For the purpose of the present analysis, reading was operationalized as reading efficiency, which combines aspects of reading accuracy and speed. The efficiency measure was computed as the number of words/pseudowords correctly read divided by the mean reaction time per item. While in previous chapters reading accuracy and speed were analysed separately, in the present chapter a combination of both measures was preferred in order to reduce the number of comparisons involved in model construction. Additionally, it is not a priori clear whether rhythmic sensitivity would be expected to impact decoding accuracy and speed differentially nor to which extent.

2.4. Rhythmic sensitivity

Non-linguistic rhythm (NLR) was assessed by tapping to a beat at three frequencies of interest: 1 Hz, 2 Hz, and 4 Hz. Stimuli were presented through headphones and responses collected digitally on a tablet. In order to increase children's motivation and engagement, the task was embedded in a narrative. In the beginning of the session, children were presented with the picture of a button and a bomb in a room on the screen (Figure 5.1).

Figure 5.1. Screenshots from the tapping to a beat task

Instructions stated: “Listen to the beat, and press the button at the same time, if you press the button at the same time, you can make the bomb blow up. Let’s see how it works”. A demo with a hand would show the correct procedure (pressing in synchrony with the beat), followed by a practice trial. If children hit the button at the same time as the beat, the bomb would enlarge until it blew up. The screen would change to black with an image of only the button, and children heard “Well done, keep pressing at the same time as the beat to go back into the house”. If they did not hit the button simultaneously in at least half of the beats, the bomb would not blow up and the audio would say “please call the research assistant for help”, and the practice trial would be repeated. This practice trial was repeated after 3 test trials—in the middle of the session. Beat frequency in the practice trial was 1.67 Hz. Practice and test trial duration was 20 seconds, and there were 2 trials for each frequency, randomly presented. One audio track per stimulus frequency was created, with the beats repeated at the desired frequency. This presentation guaranteed that inter-stimulus time would be precise and not dependent on tablet status and internal delays. Response time was recorded for each tap on the screen.

Performance was measured by computing the average time difference between each tap and its nearest stimulus for each frequency (NLR score). NLR score was computed as the average of the angular measure (a) in radians of response times for each tap computed as:

$$a = \frac{time_{dif} * 2 * pi}{k}$$

Where $time_{dif}$ is the time difference between each tap and its nearest stimulus (before or after), and k is the inter-stimulus time period in seconds (see Kirschner & Tomasello, 2009). Responses longer than 2 seconds were labelled as outliers and removed from the analysis. For each child, the best scoring trial was kept for each frequency, to account for possible fatigue effects.

2.5. Analysis rationale

First, to test the reliability of the rhythmic sensitivity task, performance was assessed through four different measures: inter-tap interval, NLR scores, correlations across frequencies and time points, and correlations with cognitive variables. Next, the contribution of rhythmic sensitivity to reading acquisition—and its mediation by phonological processing—was tested using hierarchical regression. Finally, the contribution of rhythmic sensitivity to classification accuracy of reading difficulties was tested.

3. Results

3.1. Inter-tap interval

Inter-tap interval (ITI) is defined as the mean time difference between each subsequent response, irrespective of the timing of the closest stimulus. Although it is not included in further analyses, it gives a sense of task performance. ITIs for each frequency condition and time point are shown in Figure 5.2. Although tapping frequency does not correspond to the expected frequency (1, 0.5, and 0.25 seconds respectively for each condition), they effectively change their tapping frequency for each condition. This indicates that they are trying to follow the perceived beat, although somewhat faster for the 1 and 2 Hz stimuli and slower for the 4 Hz. Secondly, as children move from K5 to G1, tapping frequency gets closer to the expected one in each condition, again suggesting that they understand the task and are trying to follow the perceived beat.

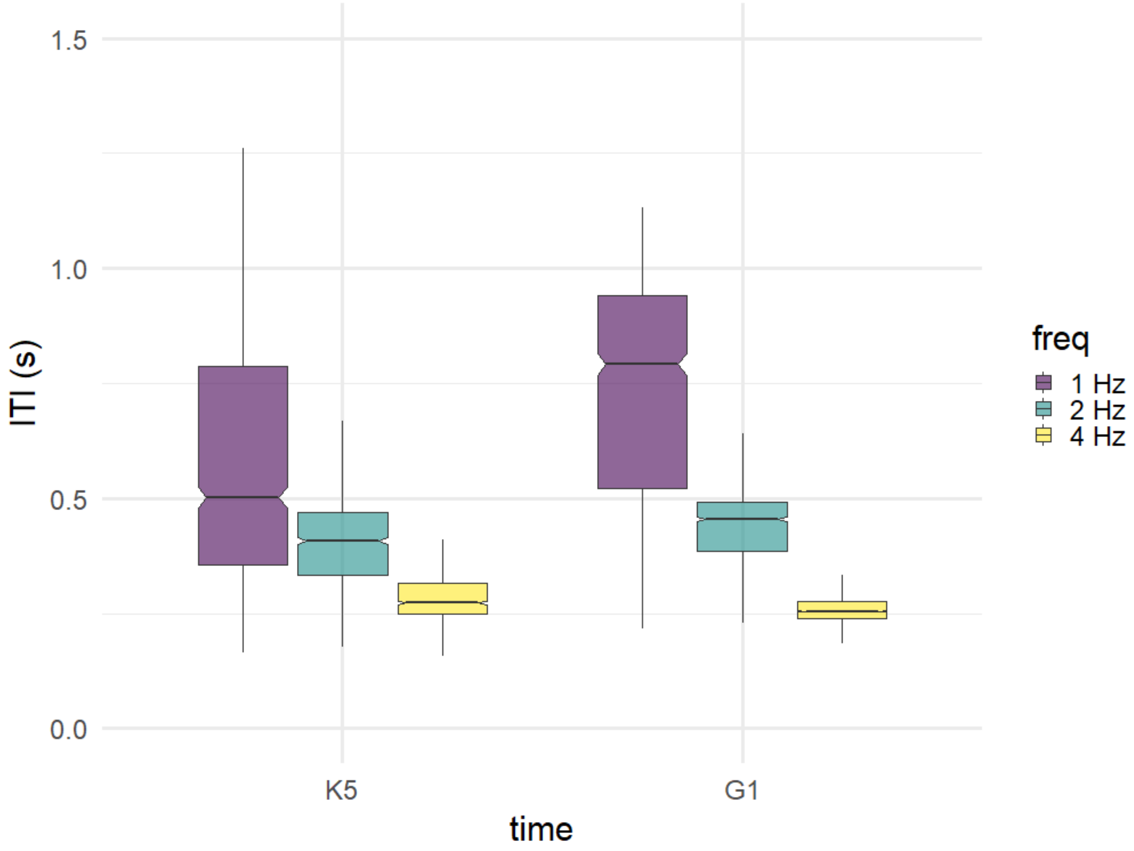


Figure 5.2. Inter-tap interval (ITI) for the three frequency conditions in kindergarten (K5) and first grade (G1). Freq: frequency.

3.2. NLR performance

Descriptive statistics for NLR by time and frequency condition are displayed in Table 5.1. Scores show large variability for all frequencies and time points and a trend towards improvement. The last column (% synch) shows the percentage of synchronizers for each time and frequency. Synchronizer class was defined through a Rayleigh test of uniformity, assessing whether there was a significant difference between the mean resultant vector length (NLR) and a uniform distribution (Kirschner & Tomasello, 2009). According to this statistic, overall performance tends to improve from K5 to G1—i.e., number of synchronizers increases with time.

Table 5.1. Descriptive statistics for NLR across frequencies and time points

time	freq	mean	sd	n	min	max	skew	% synch
K5	1	0.28	0.24	442	0.01	0.91	0.84	34.2
K5	2	0.35	0.25	442	0.00	0.93	0.48	53.6
K5	4	0.25	0.14	443	0.01	0.78	1.00	50.8
G1	1	0.44	0.27	385	0.01	0.97	-0.02	60.0
G1	2	0.48	0.26	385	0.01	0.94	-0.14	75.6
G1	4	0.27	0.17	385	0.02	0.78	1.03	55.3

Note: freq: frequency; sd: standard deviation; n: number of participants; min: minimum; max: maximum; skew: skewness; %synch: percentage of synchronizers; K5: kindergarten; G1: first grade.

A second parameter needs to be considered when interpreting performance. NLR measures the level of *consistency* in tapping behaviour, but does not account for *synchrony*—i.e., tapping phase. For example, a child might be perfectly consistent in his tapping, with an NLR of above 0.8, but be tapping counterphase to the beat. While the phase measure is not precise due to timing issues related to tablet performance, a sense of phase distribution can guide the interpretation of performance (Figure 5.3). Perfect synchronization would show a bar at 0. Synchronization performance is the best in the 1 Hz condition (children overall tend to tap in phase with the beat). At 2 and 4 Hz, synchronization performance decreases. At 2 Hz children tend to tap with a delay of approximately 0.25 cycles (about 125 milliseconds), while at 4 Hz the distribution is close to uniform. Thus, although percentage of synchronizers in K5 in the 4 Hz condition is larger than in the 1 Hz condition—which would suggest better overall performance—synchronization is poorer. Therefore, 4 Hz performance should be interpreted with caution.

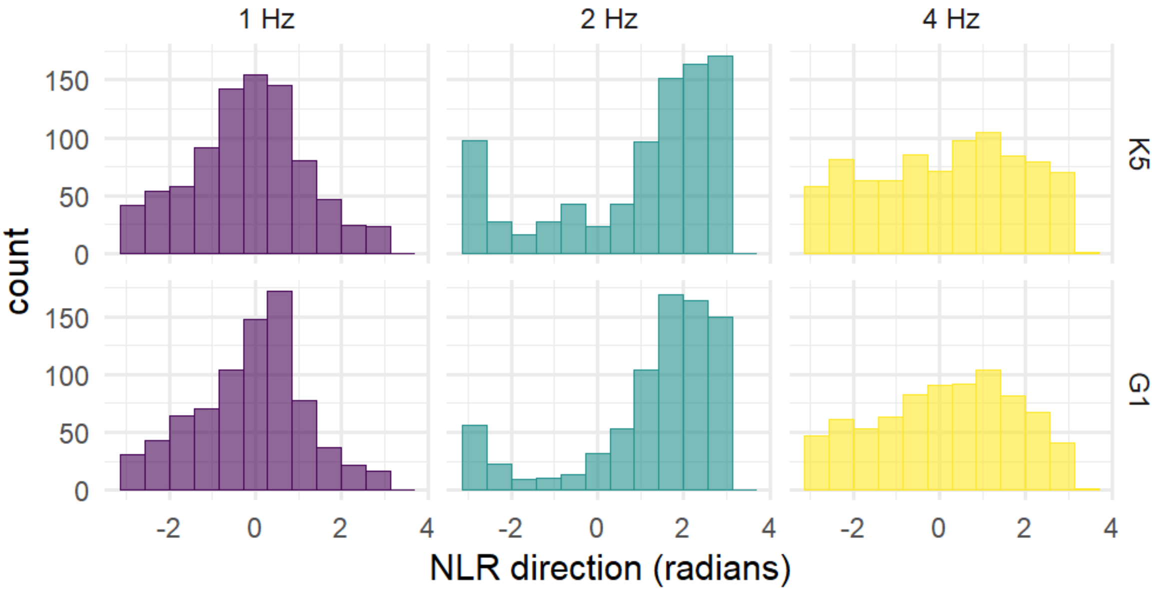


Figure 5.3. Phase distribution of tapping performance across frequencies and time points.

3.3. NLR reliability

Test-retest reliability was computed as the Pearson moment correlation between repetitions of the same trial in each time point (K5 and G1) and frequency (1, 2, and 4 Hz). Development of the skill over time was computed as the Pearson moment correlation of the best performing trial between K5 and G1 for each frequency (Table 5.2). The conditions involving 1 and 2 Hz show reliable results both in terms of test-retest and development over time. On the other hand, the 4 Hz condition shows lower reliability overall.

Table 5.2. Pearson correlation coefficients across trials and time points.

frequency	1		2		4	
	r	CI	r	CI	r	CI
K5 trials	0.567	(0.49 - 0.63)	0.590	(0.52 - 0.65)	0.228	(0.13 - 0.32)
G1 trials	0.550	(0.47 - 0.61)	0.590	(0.52 - 0.65)	0.445	(0.36 - 0.52)
K5 to G1	0.479	(0.38 - 0.56)	0.413	(0.36 - 0.52)	0.189	(0.07 - 0.30)

Note: all p values < 0.001

3.4. Correlation between NLR and other variables

The correlation structure for NLR and cognitive and reading variables across time points is shown in Table 5.3. Reported correlations represent Pearson correlation coefficients with Bonferroni corrected p values. Sample size varied according to how many children completed each task at each time point (min n: 261, max n: 328). In K5, NLR at 1 and 2 Hz moderately correlated with phonological processing (PAS and vSTM). At 4 Hz, there are no significant correlations. Regarding K5 to G1 correlations, at 1 Hz NLR moderately correlates with efficiency, phonological processing (PAS and PAP), and reading efficiency. At 2 Hz, there is a low correlation with PAP. Again, no significant correlations are observed for the 4 Hz conditions. Overall, there seems to be a robust and sustained correlation between NLR and phonological processing and reading measures across time, mainly at 1 Hz.

Table 5.3. Pearson correlation coefficients between NLR and all variables across frequencies and time points

	efficiency	PAS	IQ	vSTM	nvTSM	PAP
K5 to K5						
1	n/a	0.22*	0.06	0.25**	0.27***	0.16
2	n/a	0.25**	0.07	0.21*	0.11	0.13
4	n/a	0.11	0.02	0.18	0.06	-0.02
K5 to G1						
1	0.29***	0.33***	0.13	0.17	0.15	0.23**
2	0.15	0.17	0.07	0.07	0.02	0.21*
4	0.15	0.09	0.03	0.07	0.05	-0.03

*Note: PAS: phonological awareness at the syllable level; nvSTM: non-verbal short-term memory; vSTM: verbal short-term memory; PAP: phonological awareness at the phoneme level; * p value < 0.05; ** p value < 0.01; *** p value < 0.001.*

3.5. NLR and reading

3.5.1. Effects of frequency

To study the relationship between NLR and reading (Figure 5.4), as well as whether it is modulated by different tapping frequencies, we built a linear mixed effects model with NLR as the outcome variable and reading efficiency and the two-way interaction between reading efficiency and frequency as predictors. Random intercepts by subject where also included to account for repeated measures effects on the NLR score.

Results showed a main effect of efficiency (Type III ANOVA Wald Chi Square test, $X^2(1) = 26.45$, $p < 0.001$), a main effect of frequency ($X^2(2) = 35.8$, $p < 0.001$), and a two-way interaction ($X^2(2) = 9.44$, $p < 0.01$). Follow-up estimates using the *emmeans* package (Lenth, 2018) showed significant correlations between NLR and reading efficiency at 1 Hz (slope = 0.30, CI.low = 0.19, CI.high = 0.42, p < 0.001) and 2 Hz (slope = 0.18, CI.low = 0.06, CI.high = 0.29, $p < 0.001$), but not at 4 Hz (slope = 0.09, CI.low = - 0.03, CI.high = 0.21, $p = 0.129$). Contrasts, adjusted for multiple comparisons using Tukey's method, showed differences only between the 1 and 4 Hz conditions (slope difference = 0.21, CI.low = 0.05, CI.high = 0.38, $p < 0.007$).

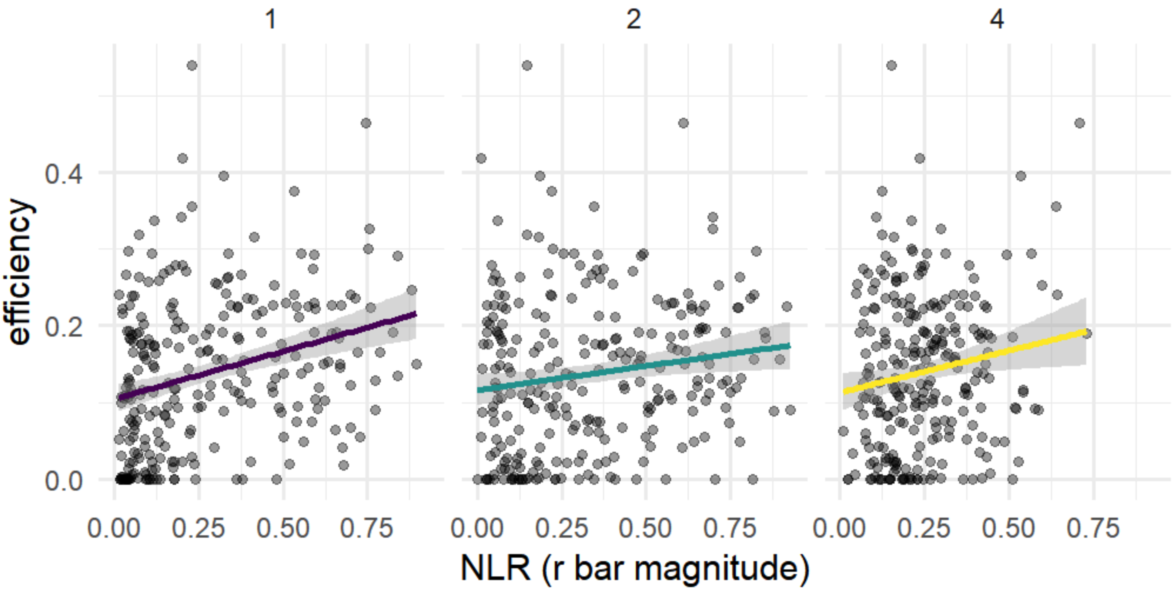


Figure 5.4. Scatter plot of K5 NLR and G1 reading efficiency scores by frequency. Regression lines represent the estimated linear trend between the two variables. Shaded areas depict standard error of the mean.

3.5.2. Effects of phonological processing

Next, in order to further study the link between NLR and reading by means of phonological processing, we took a hierarchical regression approach. Since no significant differences were observed between the NLR-reading link at 1 and 2 Hz, and no effect at 4 Hz was observed, responses were averaged over 1 and 2 Hz, and the 4 Hz condition was not analysed any further. Two linear mixed effects models were fit. In the first step, cognitive (IQ) and demographic variables (Gender and Age) were entered along with the NLR score in kindergarten to predict reading efficiency in first grade. School was included as a random intercept to account for the nesting effect of children within schools. In the next step, phonological processing (PP) measures were added (phonological awareness at the phoneme and syllable level and verbal STM). NLR coefficient significance was tested using the Satterthwaite approximation (Luke, 2017)[4]. The Nalgekerke pseudo R^2 was estimated to assess model fit through the MuMIn R package (Bartoń, 2019). The NLR coefficient in each model and the change in the NLR coefficient were inspected in order to better understand the NLR-reading link and its mediation through phonological processing.

Results show that NLR predicts reading efficiency longitudinally above and beyond the effects of cognitive and demographic variables, as well as those of phonological processing (Table 5.4). The change in the NLR coefficient in the presence of PP shows how PP mediates the effects

---

[4] Estimation of coefficient significance through likelihood-ratio tests of nested models yielded the same results.

of NLR. The effect of NLR on reading is substantially reduced when phonological processing is included in the model. In the absence of phonological processing, the effect magnitude is 0.19, meaning a one unit increase in NLR score produces almost a 0.2 unit increase in reading score. However, in the presence of PP, this effect goes down almost by half, rendering the effect of NLR on reading barely significant. Within phonological processing, verbal short-term memory seems to be capturing most of the variance previously explained by NLR.

Table 5.4. Hierarchical regression analysis predicting reading efficiency in G1 from NLR and cognitive and demographic variables in K5.

term	Beta (std)	SE	statistic	p value
Step 1				
Intercept	0.00	0.10	0.00	0.998
Age	-0.04	0.06	-0.76	0.448
IQ	0.15	0.06	2.51	0.013
Gender	0.02	0.11	0.21	0.838
nvSTM	0.31	0.06	5.04	0.000
NLR	0.19	0.06	3.14	0.002
Step 2				
Intercept	-0.03	0.10	-0.32	0.748
Age	-0.04	0.05	-0.68	0.500
IQ	0.11	0.06	2.01	0.046
Gender	0.06	0.11	0.56	0.574
nvSTM	0.22	0.06	3.47	0.001
vSTM	0.19	0.06	3.13	0.002
PAS	0.13	0.07	1.95	0.052
PAP	0.04	0.06	0.78	0.434
NLR	0.12	0.06	1.98	0.049

Note:. nvSTM: non-verbal short-term memory; vSTM: verbal short-term memory; PAS: phonological awareness at the syllable level; PAP: phonological awareness at the phoneme level; NLR: non-linguistic rhythm.

3.5.3. NLR in classification

Finally, we explored whether NLR can improve classification accuracy in a categorical model of reading efficiency. Efficiency scores were categorized in two classes with the cut-off at the first quintile of the distribution (-1.1 SD). Classification accuracy statistics (AUC, sensitivity, and specificity, see Chapter 3) were estimated for four competing models: (i) a base model including cognitive and demographic variables, (ii) the base model plus phonological processing (PP), (iii) the base model plus NLR, (iv) the full model. First, as expected, model comparisons showed that both the PP and the NLR model significantly improved model fit with respect to the base model (base vs. PP: $X^2(3) = 14.83$, $p < 0.01$, base vs. NLR: $X^2(1) = 22.85$, $p < 0.001$). Second, comparison between the PP and NLR models show no differences in model fit ($X^2(2)$

= 0, p = 1). Third, in line with the second point, sensitivity and specificity show equivalent performance in the PP and the NLR models (PP: sens = 0.9, spec = 0.74, NLR: AUC = 0.9, spec = 0.74). However, AUC shows better results in the PP than in the NLR model (PP: AUC = 0.89, NLR: AUC = 0.92). Finally, combining PP and NLR measures in one model significantly improved model fit over either the PP or the NLR model separately (PP vs. full: $X^2(1) = 16.58$, $p < 0.001$, NLR vs. full: $X^2 (3) =8.56$, $p = 0.035$). It should be noted that these were estimated on the full dataset. Thus, a better approximation of classification performance with new data should be obtained through cross-validation. In sum, equivalent classification accuracy was obtained by using rhythmic sensitivity and phonological processing measures.

4. Discussion

The main aim of this study was to test the hypothesis that rhythm sensitivity contributes to reading acquisition, and that this contribution is mediated by phonological processing. With this aim in mind, we tested rhythmic sensitivity and phonological processing at three different frequencies before the onset of reading acquisition and related it to reading efficiency at the end of first grade. Results show that rhythmic sensitivity, measured using non-linguistic stimuli, longitudinally predicts reading acquisition, both mediated by phonological processing and above and beyond it.

These findings are compatible with the framework that posits that rhythmic sensitivity aids speech segmentation and is thus involved in the development of phonological representations and, through them, in reading acquisition (Giraud & Poeppel, 2012; Goswami, 2011; Wood & Connelly, 2009). First, at all time points and reliable frequencies, rhythmic sensitivity correlated cross-sectionally with phonological processing. Secondly, the contribution of prereading NLR in longitudinally predicting reading efficiency is reduced when phonological processing measures are included in the models. Among the phonological processing variables, verbal short-term memory seems to be the main mediator, with a larger contribution than phonological awareness at the syllable and phoneme levels.

A novel aspect of this study is the association between rhythmic sensitivity and reading which is not mediated by phonological awareness or short-term memory.

On one hand, lexical stress assignment driven by morphological awareness emerge as candidate mediators (Holliman et al., 2017; Jarmulowicz, Hay, Taran, & Ethington, 2008). When words are derived, lexical stress shifts position (for example in *comunica* vs *comunicación*), therefore, morphological awareness aids in correctly assigning lexical stress. This is key in decoding polysyllabic words. In Spanish, polysyllabic words are very frequent and encountered early on.

All of the words in our reading assessment were polysyllabic, therefore correct lexical stress assignment would be a necessary skill for correct reading. This would be a particular contribution of rhythm to reading in Spanish, which would be hard to observe in early readers of English, for example, since they do not usually encounter (or are assessed on) polysyllabic words. Stress assignment is also critical in prosodic processing (i.e., prosodic stress assignment), for example in distinguishing questions from assertions in Spanish (González-Trujillo et al., 2014; A. J. Holliman et al., 2014). Thus, stress assignment can be considered a phonological skill at both the segmental (lexical and sublexical) and the suprasegmental (phrasal and sentential) levels. Therefore, the rhythm-reading relation can be considered as mediated by phonological skills in this broader sense.

At a different level of description, temporal processing has been suggested as a mediator between rhythmic sensitivity and reading acquisition (Ozernov-Palchik & Patel, 2018). This is particularly relevant for our rhythmic sensitivity task, which used a non-linguistic metrical stimulus. Our stimuli differed from speech in both its linguistic nature and its metrical structure. While our stimuli had a metrical rhythmic structure, speech has a non-metrical quasi rhythmic structure. Thus, it stands to question whether there is a common underlying mechanism in the processing of these two types of stimuli and, if so, what it is. One such candidate is the detection of temporal regularities in auditory processing (Ozernov-Palchik & Patel, 2018; A. T. Tierney & Kraus, 2013). In a study addressing this question, metrical and non-metrical non-linguistic tasks were compared in predicting letter knowledge—a precursor of reading ability—in kindergarteners. Unexpectedly, metrical rhythm was found to explain unique variance, above and beyond that non-metrical rhythm (Ozernov-Palchik et al., 2018). The results were interpreted in terms of the role that the detection of temporal regularities plays in auditory processing (Ozernov-Palchik & Patel, 2018).

To the best of our knowledge, our study is the first to systematically compare performance across frequencies. The theoretical motivation for this design was to test whether different frequencies reflected distinct sources of synchronization under the temporal sampling framework (Goswami, 2011). However, the task we used has its own limitations marked by children's motor abilities. We found that the 4 Hz condition showed low reliability and a weak correlation with other variables. It is possible that this frequency is too fast for a 5-year-old to follow, given the spontaneous tapping frequency at this age is about 2.5 Hz and the fastest forced tapping children can achieve is about 3 Hz (Drake et al., 2000). Moreover, no significant differences were observed in the rhythmic sensitivity-reading link between the 1 and 2 Hz conditions. Whether the lack of difference between frequencies reflects common mechanisms or is a limitation of the motor task needs to be explored further.

The large sample assessed in the present study allowed us to test the contribution of rhythmic sensitivity to reading before any reading experience in an unselected sample of children. This is a strength for its generalizability—as opposed to two group comparison approaches with typical readers and children with dyslexia. Moreover, it also allowed us to include a comprehensive set of covariates in the analysis while maintaining a healthy parameter/sample size ratio.

Gathering evidence from a wide variety of languages is necessary for any universal theory of reading acquisition (Goswami et al., 2014). Our results add to the scarce available evidence with prereaders in longitudinal designs (Lundetræ & Thomson, 2018) and extend them to a syllable-timed language. This is particularly relevant given the central role that prosody plays in the presented previously theoretical framework.

Finally, the current findings have implications for using non-linguistic rhythm measures as a screener for future reading difficulties. Our results show that classification accuracy improves with the inclusion of a rhythmic sensitivity measure. Crucially, this improvement is equivalent to the one obtained by phonological processing measures. For screening purposes at such a young age, tapping is a more engaging activity. Tapping, as measured through NLR, emerges then as a strong candidate for inclusion in an early screening battery.

Chapter 6

Study Four: Neural synchronization and reading acquisition

1. Introduction

Reading acquisition relies on accessing phonological representations in order to perform phoneme to grapheme correspondences (Hulme & Snowling, 2013). The development of phonological representations is contingent, among other things, on segmenting the continuous speech signal into discrete units (Cutler & Mehler, 1993). In turn, in the past two decades, neural oscillatory activity has been proposed as a mechanism underlying speech segmentation (Giraud & Poeppel, 2012). From these observations, a general framework on how neural oscillatory activity relates to reading acquisition through the development of phonological representations has been proposed (Goswami, 2011) In what follows, we describe the evidence for the role of cortical oscillations in speech processing, and its relation to reading acquisition.

1.1. Cortical oscillations and speech processing

In the brain, endogenous cortical oscillations of the electrical activity of ensembles of neurons serve a range of functions in cognitive processing, from sustained attention to memory to visual processing (Haegens & Zion Golumbic, 2018). In sensory processing, rhythmic neural oscillations have been shown to serve an “active sensing” function (Schroeder & Lakatos, 2010), where its phase or amplitude is modulated by external visual or auditory stimuli. Interestingly, both the speech signal and neural ensembles in the auditory cortex oscillate at very similar frequencies (Poeppel & Assaneo, 2020). More than a coincidence, these arise probably as the result of the rhythmic movement of the jaw during speech production (Poeppel & Assaneo, 2020). When decomposing the speech signal into frequency bands, it is possible to observe energy peaks mainly at three frequencies: delta (0.5 – 3 Hz), theta (4 – 8 Hz), and gamma (above 30 Hz). In speech, these frequencies broadly represent the change rate corresponding to prosodic, syllabic, and phonemic information, respectively. A similar pattern is observed for neural activity in the auditory cortex, with fluctuations at delta, theta and gamma bands (Gross et al., 2013).

Crucially, cortical auditory oscillations have been shown to synchronize to incoming auditory input. A seminal work by Gross and colleagues studied cortical oscillations in response to speech. They found that cortical oscillations *synchronize*[5] to speech in a

---

[5] They actually use the term entrainment. Entrainment implies that one signal becomes coupled to another signal. Synchronization is a more general term in which two signals are aligned. The main difference stems from one signal *driving* another, versus each of them oscillating — in synchrony, but— independently. Since the term entrainment is often used without actually

hierarchical manner (Gross et al., 2013). First, they found speech-brain coupling. They showed that, in response to edges in the speech envelope, delta and theta oscillations modulate their phase (phase-reset) and gamma oscillations modulate their amplitude, which results in increased synchronization between brain and speech. Second, they found brain-brain coupling. They showed that theta and gamma cortical oscillations are nested, and that this nesting works by phase-amplitude coupling, where the phase of theta oscillations modulates the amplitude of gamma oscillations in the auditory cortex. Crucially, these two forms of coupling were stronger when participants were listening to a story than when listening to the same story played backwards, underscoring its role in linguistic processing rather than being purely acoustic in nature.

Approximately at the same time, Anne Lise Giraud and David Poeppel (2012) proposed a framework in which this hierarchical organization of cortical oscillations subserves speech segmentation in the auditory cortex. By analysing available evidence and building a computational model of speech processing, they suggested that theta-gamma coupling in auditory cortex allows for analysing the speech signal into two distinct speech relevant temporal time scales, and that although occurring in parallel, these two time scales remain bounded by their coupling. Crucially, given its characteristic frequencies, gamma and theta would underly syllabic and phonemic processing.

1.2. Cortical oscillations and reading in older children and adults

In the reading domain, based on similar evidence, Usha Goswami proposed a temporal sampling framework (TSF) for developmental dyslexia. Under this framework, poor synchronization between cortical oscillations in the auditory cortex and the speech input could explain phonological deficits observed in children with dyslexia. On one hand, poor synchronization at theta rate could affect syllable level representations and prosodic structure. On the other, poor synchronization at gamma rate could affect phonemic level representations. Either of which would result in deficient phonological representation and thus in reading difficulties. Goswami proposed that poor synchronization at both frequencies could be a consequence of poor sensitivity to rise time in speech envelope. Rise time refers to the acoustic instantiation of “speech edges”, that is, salient peaks in the speech envelope that result from the rapid changes of energy corresponding to the onset of stressed syllables.

---

proving entrainment, we will prefer the term *synchronization* to refer to both entrainment in the narrow and in the broad sense (Obleser & Kayser, 2019).

With this framework in mind, many studies have assessed differences in neural synchronization between dyslexics and typical readers. In adults, studies have shown differences between dyslexics and controls in delta (Hämäläinen et al., 2012), theta (Lizarazu, Lalllier, et al., 2015) and gamma (Lehongre et al., 2011; Lizarazu, Lalllier, et al., 2015), using linguistic or non-linguistic stimuli. In children, differences in neural synchronization have also been observed in delta (Cutini et al., 2016; Lallier et al., 2016; Power et al., 2013, 2016), theta (Lizarazu, Lalllier, et al., 2015) and gamma (Lehongre et al., 2011). However, these studies stem from diverse languages, populations, stimuli, techniques, and methods, which makes it hard to integrate them. For example, at delta rate, one study found dyslexics showed weaker synchronization than controls (Molinaro et al., 2016), another study found dyslexics showed stronger synchronization than controls (Cutini et al., 2016), whereas a third one found no differences (Lizarazu, Lalllier, et al., 2015). Additionally, a recent study trying to replicate the overall previous findings mostly failed to do so, except for finding weaker delta and gamma synchronization in dyslexics vs. controls (Lizarazu et al., 2021).

Increasing the complexity of the situation, almost all studies were comparing dyslexic and typical readers. Therefore, any observed difference between groups might represent a consequence of the reduced reading experience of dyslexics rather than a cause (Huettig et al., 2018), since it is now well established that reading experience modifies speech processing (Castles & Coltheart, 2004). If synchronization in cortical oscillations underlies speech segmentation and the formation of phonological representations, we would expect to observe these effects before any reading experience.

1.3. Cortical oscillations and reading in prereaders

To the best of our knowledge, only three studies up to date have explored the neural response to the temporal properties of auditory stimuli (and its relation to reading) in prereaders. Woodruff-Carr et al (2014) examined neural synchronization to /ba/, /da/, and /ga/ syllables presented at 4.5 Hz and found that the precision in neural synchronization correlated with preliteracy skills (phonological short-term memory and phonological awareness). De Vos et al. (2017) examined neural synchronization to amplitude modulated noise at speech relevant rates (delta at 4 Hz, and beta at 20 Hz) longitudinally in children with or without family risk for dyslexia. They found increased neural synchronization for dyslexic children in the beta range only after children started reading, but not before. A more recent report on this same sample found reduced subcortical neural synchronization at high gamma (80 Hz) in dyslexic children with family risk for dyslexia with respect to typical readers without family risk (De Vos et al., 2020).

Moreover, Rios-Lopez et al. (2020) carried out a longitudinal analysis of children from 4 to 6 years of age, using continuous speech stimuli. Their sample was composed of typically developing children. They analysed neural responses in the 0 – 10 Hz range and found neural synchronization to delta rate only (0.5 Hz). They also found that delta synchronization correlated with children's responses to comprehension questions, suggesting a role for delta synchronization in general linguistic abilities (although it could also relate to attention to the speech stream). In sum, in prereaders, links between neural synchronization and preliteracy skills or reading have been shown at theta (Woodruff Carr et al., 2014), at high gamma but not at beta or theta (De Vos et al., 2017, notice the effects observed in beta were only observed after reading acquistion, 2020), and at delta but not theta (Ríos-López et al., 2020). Moreover, the three studies used different stimuli and analysis making them hard to compare. For example, it is not clear whether the differences in results between studies stems from using linguistic (Ríos-López et al., 2020) vs. non-linguistic stimuli (De Vos et al., 2017), or whether they arise from looking at synchronization through power (De Vos et al., 2017) or coherence analysis (Ríos-López et al., 2020). In sum, while there is some evidence that neural synchronization in prereaders relates to future reading acquisition, it is not clear at which frequency bands the effect is observed, and to what extent it is modulated by the linguistic nature of the stimulus.

In the present study we aimed at contributing to the understanding of the role that neural synchronization plays in reading acquisition by providing novel evidence from a longitudinal approach. We examined neural synchronization in prereading children at midterm kindergarten, and their reading development one year later after they had received reading instruction. Our hypothesis was that neural synchronization at theta and/or delta tested in prereaders would correlate to reading acquisition one year later, in line with the temporal sampling framework (Goswami, 2011) and recent evidence (De Vos et al., 2017; Ríos-López et al., 2020; Woodruff Carr et al., 2014) For this purpose, we computed neural synchronization in response to modulated white noise at speech relevant rates (delta, theta and an 8 Hz control condition) in a non-linguistic stimulus (amplitude modulated noise), and related it to reading skills one year later.

2. Methods

2.1. Sample

Forty children attending kindergarten (K5) took part in the study (21 males, age range 5 – 6.5 years, mean = 6.1). All parents of participants provided informed written consent and all children verbally agreed to participate. All participants were Spanish native speakers, had normal or corrected-to-normal vision and reported no hearing impairments. Behavioural data was collected between June and August during kindergarten year (mid-term), and between October and December while in first grade (end of term). Electrophysiological data was collected between November and February during kindergarten year (end of term). Data from four children was discarded due to noisy signal (two children), technical issues during recording (one child) and falling asleep during recording (one child). The final sample was composed of 36 children.

2.2. Behavioural measures

During kindergarten, decoding and IQ were assessed. Decoding was assessed through a list of 15 frequent words and 15 pseudowords presented in paper. Nonverbal IQ was measured using the Matrix Reasoning subtest of the Spanish version of the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 2001). At the end of first grade, decoding accuracy was assessed by presenting a list of 30 words and 30 pseudowords digitally, one word per screen. For a full description of the measures please refer to Chapter 2.

2.3. Neural measures

2.3.1. Stimuli

Stimuli consisted of amplitude modulated white noise. Modulation frequency was 2, 4, 8 and 60 Hz with 100% depth and a non-modulated condition as in Lizarazu et al. (Lizarazu, Lallier, et al., 2015). Each condition was presented in 10 second trials, repeated 24 times. There was no inter-trial-interval. Order of presentation was random.

Stimuli were presented binaurally through Etymotic ETY Kids 5 insert earphones. Sound pressure level adjustment for each child consisted in listening to a recorded sentence (*Donde viven los monstruos* [Where monsters live]) and repeating it correctly. During the whole session children viewed silent cartoons displayed in a projector on the wall in order to maintain them entertained and as quiet as possible. Session was interrupted if children showed signs of boredom or tiredness.

Presentation was coded in PsychoPy (Peirce, 2008), using the sound library with *pyo* backend in Windows 7. Given that performance issues have been reported for *pyo* in Windows OS, following conclusion of data collection, we studied the delay between the timing of the sound trigger and the actual sound output using in-house developed hardware and software. Analysis of 200 trials revealed an average 190 ms delay with a non-negligible standard deviation (SD = 27 ms). This variable delay would likely affect estimates of phase synchrony between stimuli and response. Thus, analysis was focused on power estimates.

2.3.2. EEG recording and processing

Recording. EEG data was acquired using a Biosemi Active Two system, with 32 electrodes in a 20-10 layout. Activity was referenced online to the common mode sense (CMS, active electrode) and grounded to a passive electrode (Driven Right Leg, DRL). Data was digitized at 512 Hz.

Pre-processing. EEG signal was processed using Fieldtrip toolbox (Oostenveld et al., 2011) in Matlab R2018a (The Mathworks Inc., 2018) and custom developed code. In the pre-processing step, the continuous EEG signal was band passed with a two-pass fourth-order Butterworth filter between 0.1 and 40 Hz, baselined to 0.8 s prior to stimulus onset and re-referenced to Cz electrode. In each 10 second trial, the first second was discarded due to an observed increase in noise in response to the incoming stimuli, and was redefined into 2 second epochs with 1 second overlap for all conditions. On each 2 second epoch, artifact rejection was based on an adaptation of Junghöfer et al (2000). Channels and epochs were rejected if, for each stimulus condition, they surpassed a threshold defined by the median of channels/trials standard deviation according to following equation (see Flo, 2019):

$$th = mstd + 2 \times \sqrt{\frac{\sum_i^N (std_i - mstd)^2}{N}} \tag{1}$$

Rejected channels were interpolated using the default "weighted" method in the ft_channelrepair function and a Biosemi 32 template for defining neighbours with the ft_prepare_neighbours function. According to the template used for neighbour selection, each channel has on average 6 neighbours, ranging from 3 to 8 according to channel position. On average, after artefact rejection, number of epochs per child per condition was 140.

Next, power estimates were obtained through the discrete Fourier transform via the fft Matlab function. For each condition, 2-second-epochs were concatenated in groups of 5 in order to increase spectral resolution. Number of epochs was standardized across children by limiting it to a range between 90 and 120; for children with more than 120 epochs, subsequent epochs were discarded. Thus, for each child there were between 18 and 24 sweeps per condition. Next, sweeps were averaged in the time domain and transformed into the frequency domain. Signal was padded with zeros to the next power of two in order to improve performance.

SNR. The previous processing steps yielded power estimates per child per channel per condition with a spectral resolution of 0.0625 Hz (

Figure 6.1). Signal-to-noise ratio was used to quantify the degree of synchronized neural activity (De Vos et al., 2020). SNR was computed for each stimulus frequency as:

$$SNR_f(dB) = 10 \times log_{10}\left(\frac{P_f}{P_{f\pm 10bins}}\right) \qquad (2)$$

where P_f is the response power, and $P_{f\pm 10bins}$ is the power in 10 adjacent bins from the stimulus frequency.

Next, in order to obtain the specific response to the stimulus of interest, relative SNR was defined as the subtraction of the SNR for each stimulus frequency from the SNR from unmodulated stimuli (control condition). For each AMN frequency, 3 relative SNR were computed corresponding to the 3 stimulus frequencies. This resulted in a 3 x 3 design of AMN frequency (2, 4 and 8 Hz) x relSNR frequency (2, 4 and 8 Hz).

$$relSNR_f = SNR_f - SNR_{ctrl} \qquad (3)$$

Neural synchronization was defined when the response to a stimulus of interest was significantly larger that the response observed for the control condition, i.e., a relSNR significantly larger than zero.

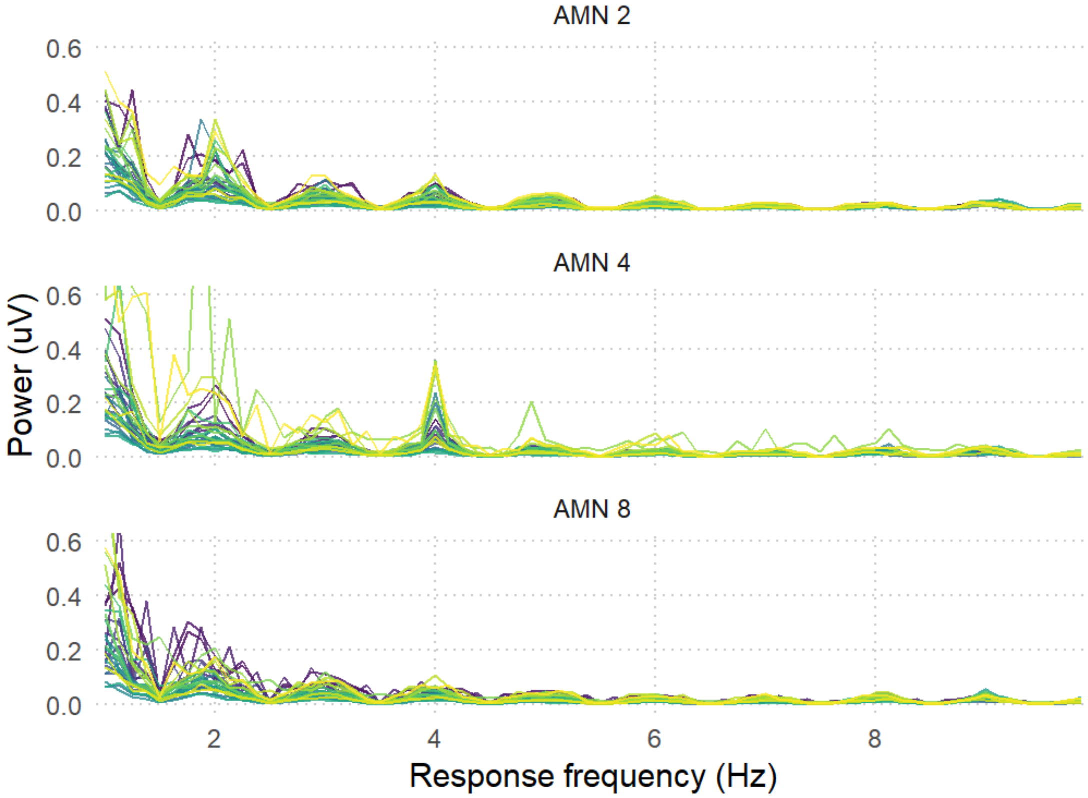


Figure 6.1. Power spectrum for each stimulus frequency (top: 2 Hz; middle: 4 Hz; bottom: 8 Hz). Each line represents one electrode. AMN: amplitude modulated noise.

2.4. Statistical analysis

Linear mixed effects models were built to predict brain responses via the *lme4* package in R (Bates et al., 2015; R Core Team, 2018). For all models, planned comparisons were obtained through the *emmeans* package. Degrees of freedom were estimated through the Satterthwaite method, and p-values were adjusted for multiple comparisons through the false discovery rate (fdr) method.

3. Results

First, neural synchronization to the stimulus conditions for each response frequency were examined. Next, the topographical distribution of the response was investigated. In order to avoid ad-hoc electrode selection criteria, neural synchronization was examined in an overarching model with all electrodes. Finally, relation between neural synchronization and reading acquisition was investigated.

3.1. Neural synchronization to AMN

A linear mixed effect model was computed with relSNR as outcome and main effects of stimulus frequency, response frequency and their interaction. Random intercepts by subject were included in order to account for the repeated measures of electrodes over subjects.

Model results showed main effects of stimulus frequency, of response frequency and a significant interaction (all $p < 0.001$). Estimated marginal means and planned comparisons were obtained for each stimulus and response frequencies for all main effects and the interaction (Table 6.1 and Figure 6.2). An estimated marginal mean for relSNR larger than zero was interpreted as a specific neural synchronization to the presented stimuli (i.e., significantly larger neural synchronization to AMN than to unmodulated noise stimuli).

Table 6.1. Estimated marginal means and its contrasts for each stimulus and response frequencies

			estimated marginal means				
stim	resp	estimate	df	conf.low	conf.high	statistic	p.value
AMN2	-	0.90	40.92	0.13	1.68	2.35	0.024
AMN4	-	1.89	40.92	1.11	2.66	4.91	0.000
AMN8	-	0.25	40.92	-0.52	1.03	0.66	0.516
-	SNR2	1.29	40.92	0.51	2.06	3.36	0.002
-	SNR4	2.05	40.92	1.27	2.82	5.33	0.000
-	SNR8	-0.30	40.92	-1.07	0.48	-0.77	0.446
AMN2	SNR2	1.97	61.43	1.12	2.82	4.64	0.000
	SNR4	1.14	61.43	0.29	1.99	2.69	0.009
	SNR8	-0.41	61.43	-1.26	0.44	-0.96	0.342
AMN4	SNR2	1.23	61.43	0.38	2.08	2.90	0.005
	SNR4	4.56	61.43	3.71	5.41	10.74	0.000
	SNR8	-0.14	61.43	-0.99	0.71	-0.33	0.744
AMN8	SNR2	0.66	61.43	-0.19	1.51	1.56	0.124
	SNR4	0.43	61.43	-0.42	1.28	1.01	0.315
	SNR8	-0.34	61.43	-1.19	0.51	-0.80	0.428
			contrasts				
AMN2 - AMN4	-	-0.98	10000	-1.41	-0.56	-5.40	0.00
AMN2 - AMN8	-	0.65	10000	0.22	1.08	3.57	0.00
AMN4 - AMN8	-	1.63	10000	1.21	2.06	8.97	0.00
-	SNR2 - SNR4	-0.76	10000	-1.18	-0.33	-4.15	0.00
-	SNR2 - SNR8	1.58	10000	1.16	2.01	8.69	0.00
-	SNR4 - SNR8	2.34	10000	1.91	2.77	12.85	0.00

AMN2	SNR2 - SNR4	0.83	10000	0.09	1.57	2.62	0.02
	SNR2 - SNR8	2.38	10000	1.64	3.12	7.53	0.00
	SNR4 - SNR8	1.55	10000	0.81	2.29	4.91	0.00
AMN4	SNR2 - SNR4	-3.33	10000	-4.07	-2.59	-10.55	0.00
	SNR2 - SNR8	1.37	10000	0.63	2.11	4.35	0.00
	SNR4 - SNR8	4.70	10000	3.96	5.44	14.90	0.00
AMN8	SNR2 - SNR4	0.23	10000	-0.51	0.97	0.73	0.74
	SNR2 - SNR8	1.00	10000	0.26	1.74	3.17	0.00
	SNR4 - SNR8	0.77	10000	0.03	1.51	2.44	0.04

Note: Degrees of freedom method used: Satterthwaite; conf.low: confidence interval lower bound; conf.high: confidence interval upper bound; confidence interval level: 95%; p.value: p value; adjustment: Tukey.

With respect to *stimulus* frequencies, neural synchronization was found for 2 and 4 Hz stimuli but not for 8 Hz. Neural synchronization was different among all pairwise comparisons, with responses for the 4Hz stimuli being largest, followed by responses to the 2 Hz and 8 Hz stimuli (Table 6.1). The same pattern was observed with respect to *response* frequencies: neural synchronization was found for 2 and 4 Hz responses, but not for 8 Hz, and was different among all pairwise comparisons, with 4Hz responses being largest, followed by 2 Hz and 8 Hz.

When looking at the interaction between *stimulus and response* frequencies, results showed that, in each stimulus frequency, responses were largest when stimuli and response frequencies coincided. In other words, when children heard 4 Hz stimuli, they showed larger responses at 4 Hz than at 2 or 8 Hz; and when they heard 2 Hz stimuli, they showed larger responses at 2 Hz than at 4 or 8 Hz. This was not the case in the 8 Hz stimuli, where none of the responses were significantly different from the control condition.

Thus, results showed that brain responses tune to auditory stimuli at frequencies relevant to the speech input (2 and 4 Hz). Furthermore, neural synchronization was largest for response frequencies corresponding to the stimulus frequency. In order to investigate the topographical distribution of neural synchronization, we further explored neural synchronization by electrode.

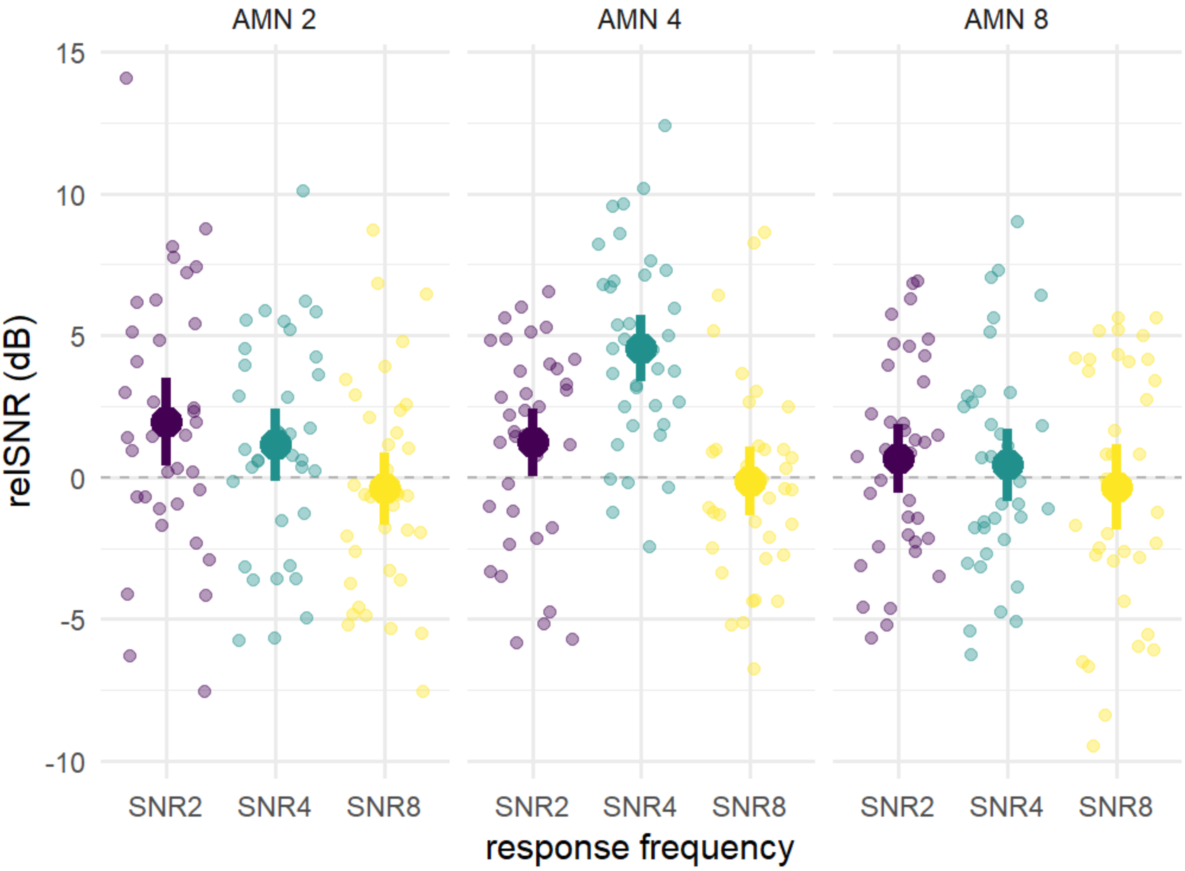


Figure 6.2. Relative SNR responses averaged across electrodes for each stimuli and response frequencies. Dots represent each child's individual average across electrodes. AMN: amplitude modulated noise; SNR: signal to noise ratio.

3.2. Topography

The topographical distribution of neural synchronization was studied for delta and theta stimuli (2 and 4 Hz, respectively). Given the previous results, responses at 2 Hz were analysed for delta stimuli, and responses at 4 Hz were analysed for theta stimuli.

Two linear mixed effects models were fit (one for each frequency) with neural synchronization as outcome and electrode as predictor. Random intercepts for children were included to account for the repeated measures of electrodes. Both frequencies showed a significant effect of electrode (delta: $X^2(30) = 49.4$, $p = 0.014$; theta: $X^2(30) = 81.7$, $p < 0.001$). T-tests for each electrode were computed, with false-discovery rate correction for multiple comparisons. Model estimates of brain responses at each electrode for delta and theta rates are displayed in Figure 6.3.

For delta, two clusters of neural synchronization were observed, one anterior (AF4, F7, F4, FC1, Fz) and one posterior (PO4, O1, O2, Pz, Oz). A significant F8 response was also observed, but SNR was negative, meaning the response to the unmodulated stimuli was larger than to the modulated stimuli. Whether this could be interpreted as an artifact, given the bizarre and unique pattern, is not clear. Once false discovery rate correction was applied for multiple comparisons, synchronization is observed at one anterior site only (F7).

For theta, a broad scalp-wide neural synchronization is observed, except for four central electrodes (FC1, FC2, C4 and CP1, FDR corrected).

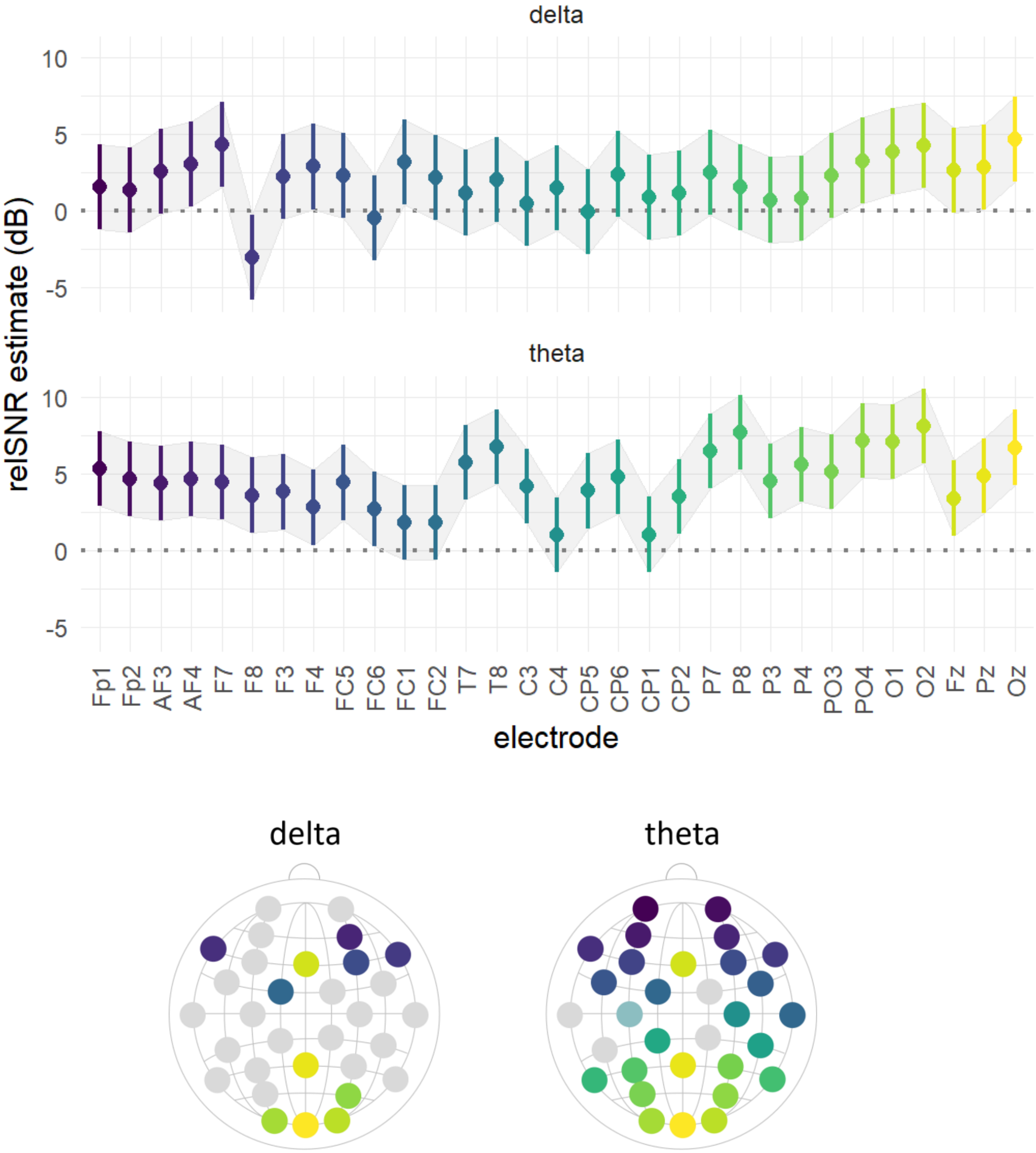


Figure 6.3. Top. Neural synchronization for each frequency and electrode. Dots represent the estimated marginal mean of the relative SNR response across children, and its 95% confidence interval. Grey shaded area corresponds to 95% confidence intervals for each electrode. Electrodes are arranged left to right from anterior to posterior. Midline electrodes correspond to the last three columns. Bottom. Topographical distribution of neural synchronization.

Coloured points represent electrodes with significant synchronization (not corrected for multiple comparisons). For delta, two clusters of neural synchronization can be observed, one anterior and one posterior. For theta, a scalp-wide neural synchronization is observed. AMN: amplitude modulated noise, relSNR: relative signal-to-noise ratio.

3.3. Reading

Next, we tested how neural synchronization relates to reading acquisition at each frequency. During kindergarten, most children (n = 33) could not read any of the presented words and pseudowords, and three children were already reading (they could read 9 out of 15 pseudowords on average). At the end of first grade, reading scores displayed a bimodal distribution. Two groups were defined[6] based on a -1 z score threshold into *poor readers* (PR, mean accuracy = 0.13, min = 0, max = 0.45,n = 7) and *typical readers* (TR, mean accuracy = 0.89, min = 0.7, max = 1, n = 28).

We computed two linear mixed effects regression models —one for delta and one for theta— with neural synchronization as outcome, and fixed effects for reading, electrode (delta: n = 4; theta: n = 27) and its interaction. Random intercepts for children were included to account for the repeated measures of electrodes within children. Levene’s test was used to check the homogeneity of variance assumption, which was met for both models (delta: F(1,142) = 0.002, p = 0.96); theta: F(1,970) = 1.03 p = 0.96 (Foster et al., 2011)).

Table 6.2. Delta neural synchronization at each electrode and frequency, and contrasts between reading groups

estimated marginal means						
group	elec	relSNR estimate	statistic	conf.low	conf.high	p.value
poor reader	F7	-1.88	-0.65	-8.50	4.74	0.519
typical reader		5.81	4.08	2.56	9.07	0.000
poor reader	O1	3.71	1.28	-2.91	10.33	0.204
typical reader		3.92	2.75	0.67	7.17	0.014
poor reader	O2	3.88	1.34	-2.73	10.50	0.184
typical reader		4.34	3.05	1.09	7.60	0.006
poor reader	Oz	6.09	2.10	-0.53	12.71	0.039
typical reader		4.30	3.01	1.04	7.55	0.007
contrasts						
poor reader - typical reader	F7	-7.69	-2.38	-14.12	-1.27	0.02
poor reader - typical reader	O1	-0.21	-0.06	-6.63	6.22	0.95

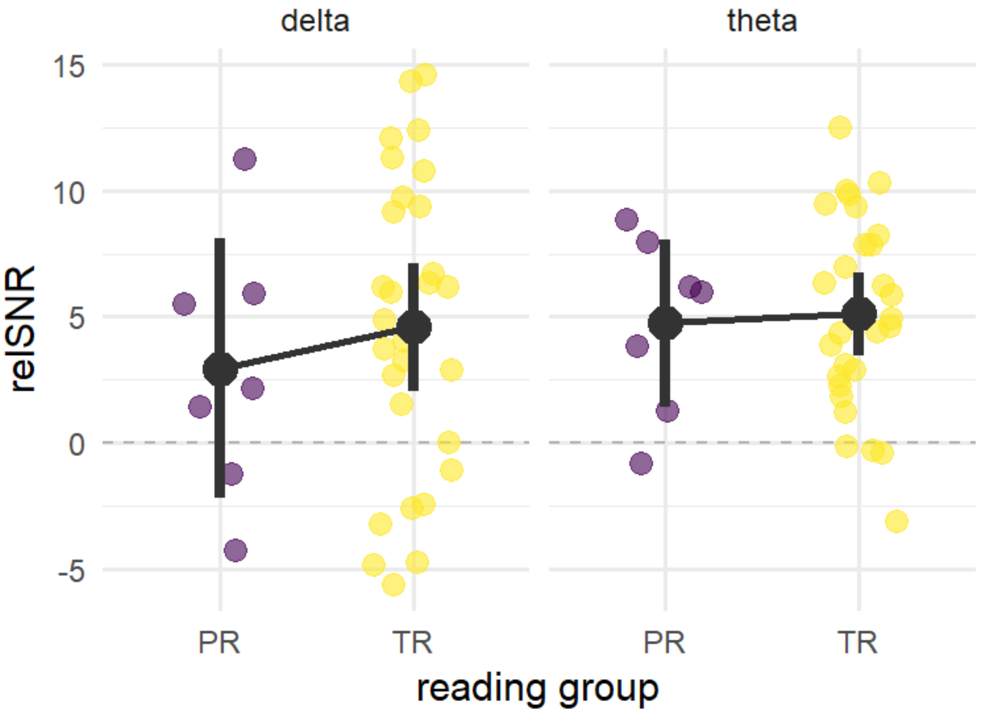
---

[6] Equivalent results were found for a continuous model, and by using fluency instead of decoding.

poor reader - typical reader	O2	-0.46	-0.14	-6.89	5.96	0.89
poor reader - typical reader	Oz	1.79	0.56	-4.63	8.22	0.58

Note: elec: electrode; relSNR estimate: relative SNR estimate (in decibels); conf.low: 95% confidence interval lower bound; conf.high: 95% confidence interval upper bound; statistic: Student's t statistic. P values are FDR adjusted.

At delta, a main effect of reading ($X^2(1) = 5.66$, $p = 0.017$), a marginal main effect of electrode ($X^2(3) = 7.29$, $p = 0.063$), and their interaction was observed ($X^2(3) = 8.79$, $p = 0.032$). Post-hoc tests showed that poor readers did not show neural synchronization (mean relSNR = 2.95, SE = 2.2, 95% CI = [-2.22, 8.12], t = 1.33, $p_{FDR} = 0.18$), but typical readers did (mean relSNR = 4.59, SE = 1.08, 95% CI = [2.06, 7.13], t = 4.24, $p_{FDR} = 0.0003$) (Figure 6.4). Overall contrasts did not show differences between groups (PR – TR estimate = -1.64, SE = 2.46, 95% CI = [-6.63, 3.35], t = -0.67, $p_{FDR} = 0.50$). However, when electrodes were included, significant differences between groups were observed at the F7 electrode but not in the occipital cluster[7] (Table 6.2). A linear model predicting neural synchronization at F7 from reading group, Age, and IQ, showed that reading group predicts neural synchronization above and beyond Age and IQ (model F(3,32) = 2.83, p = 0.053; reading group coefficient: estimate = 7.89, SE = 3, t = 2.65, p= 0.013).



---

[7] Non-parametric statistics (Mann-Whitney test) of typical vs. poor readers relSNR at the F7 electrode yielded equivalent results (W = 53, p = 0.053).

Figure 6.4. Neural synchronization by reading group at delta and theta rates. Error bars indicate 95% confidence intervals. PR: poor reader; TR: typical reader. Dots represent individual children (averaged across electrodes).

At theta, no significant main effects or interactions were observed. Post-hoc tests confirmed that both reading groups showed neural synchronization to the stimulus (PR: mean relSNR = 4.74, SE = 1.42, 95% CI = [1.41, 8.07], t = 3.34, p_{FDR} = 0.002; TR: mean relSNR = 5.10, SE = 0.70, 95% CI = [3.46, 6.73], t = 7.32, $p_{FDR} < 0.0001$), and contrasts confirmed no differences between groups (t = -0.22, p = 0.82) .

4. Discussion

4.1. Neural synchronization in prereaders

Our findings show that prereading children show neural synchronization both at delta and theta rates (but not at 8 Hz), thus showing a cerebral specialization for auditory processing at speech-relevant rates. These results confirm findings reported in older children and adults (Lizarazu et al., 2021) and add novel evidence to the scarce available data coming from prereaders. Previous studies with prereaders have reported neural synchronization to theta rate (Vanvooren et al., 2014) and delta rate (Ríos-López et al., 2020), but none of the previous studies have found neural synchronization to both rates. While Vanvooren et al did not test delta rate, Rios-López tested both frequencies in response to speech and failed to find neural synchronization at theta. Very recently, novel evidence from infant studies has also shown synchronization at delta (Attaheri et al., 2020) and theta (Ortiz Barajas et al., 2021) in response to infant-directed speech. Thus, the emerging picture suggests that that neural synchronization to auditory stimuli at speech-relevant rates develops very early on, possibly, in a continuous manner.

Comparing neural synchronization to both frequencies allowed us to show that neural synchronization was stronger and more widely distributed for theta than to delta rate, at least for non-linguistic stimuli. The topographical distribution of the observed effects warrants discussion.

For theta synchronization, the large scalp-wide distribution was somewhat surprising. Previous studies have reported theta synchronization at temporo-parietal electrodes with fNIRS (Cutini et al., 2016), EEG (Lehongre et al., 2013; Vanvooren et al., 2014) and MEG (Lizarazu, Lalllier, et al., 2015), broadly corresponding to auditory cortex in the temporal lobe. However, source reconstruction of auditory steady state responses to a 40 Hz AMN stimuli have found sources both in central auditory pathways —both cortical and subcortical, including brainstem and primary auditory cortex— and extra auditory pathways broadly distributed —including pre and post central gyri, orbitofrontal, parahippocampal, occipital, superior parietal and cingulate gyri—(Farahani et al., 2020). When the stimulus was limited to the theta range, sources have been found in the frontal lobe and medial limbic structures (Farahani et al., 2017), and in associative auditory and non-auditory cortex (Giraud et al., 2000). Thus, it is possible that the observed topography results from our data-driven approach to electrode selection, and that it originates from both auditory and extra-auditory sources.

For delta synchronization, we found a frontal and an occipital cluster, as opposed to the expected temporo-parietal distribution (Lehongre et al., 2013; Lizarazu, Lallier, et al., 2015; Ríos-López et al., 2020). However, delta synchronization has also been described in both auditory and non-auditory cortices. On one hand, delta synchronization has been observed in the auditory cortex, involved in bottom-up segmentation of the speech input (Ghitza, 2017; Lizarazu, Lalllier, et al., 2015; Molinaro et al., 2016). On the other hand, delta synchronization has also been shown in the frontal lobe —in particular IFG and precentral gyrus— exerting top-down modulations on delta and theta activity of the auditory cortex (Molinaro et al., 2016; Park et al., 2015). These top-down modulations have been involved in the grouping of words into syntactic phrases (Ding et al., 2015; Meyer et al., 2019), in temporal predictions during speech processing (Rimmele et al., 2018), and in sensory chunking of articulated sounds (Boucher et al., 2017). The observed distribution in the current data is more compatible with top-down frontal effects than bottom-up temporal processing. Moreover, since our stimuli were non-linguistic, it is more likely a reflection of a non-linguistic mechanisms such as temporal prediction or sensory chunking. Importantly, we failed to find neural delta synchronization at the auditory level. At this time, it is hard to tell whether this reflects developmental differences (although see Ríos-López et al., 2020) or if it is a consequence of our paradigm or processing.

4.2. Neural synchronization and reading

With respect to neural synchronization and reading, this is the first study to show that differences in delta —but not theta— synchronization, precede and explain future reading acquisition. The results provide partial support for the temporal sampling framework (Goswami, 2011) in that the quality of neural synchronization at low frequencies affects later reading acquisition. In particular, it adds novel evidence in the differential role played by delta vs. theta synchronization in reading acquisition. Our findings suggest that, while prereaders show neural synchronization at both delta and theta rates, only synchronization at delta rate relates to reading acquisition. It is worth noting that we did not directly test in a single model the contribution of delta vs theta in explaining the relation between reading and neural synchronization. However, the characteristics of theta synchronization suggest that the lack of a role for theta in this relation does not stem from lack of power or methodological issues. Theta synchronization is both stronger and much more widely distributed than delta synchronization, suggesting that if there indeed was an effect between theta synchronization and reading, we would have been able to detect it. Importantly, at theta rate, both reading groups showed significant neural synchronization (that is, they showed significantly larger neural synchronization for amplitude modulated white noise than for unmodulated white noise). Thus, it seems that synchronization at theta rate, although present in prereaders, it is not particularly relevant for reading acquisition. These results contradict previous findings in older children and adults which showed significant differences between dyslexics and controls at theta rate (Lizarazu, Lalllier, et al., 2015) but support several studies in older children and adults reporting no differences between groups (De Vos et al., 2017; Hämäläinen et al., 2012; Lehongre et al., 2013; Lizarazu et al., 2021; Power et al., 2016). Moreover, available evidence shows that theta synchronization mainly reflects acoustic (vs. linguistic) processing of the input. For example, Boucher et al. (2017) showed equivalent theta synchronization in processing tones, nonsense syllables or utterances, and Molinaro and Lizarazu also found equivalent theta synchronization in speech processing, vs. rotated speech or amplitude modulated noise (Molinaro & Lizarazu, 2017). These and additional studies underscore theta's role in acoustic/perceptual processing (Etard & Reichenbach, 2019; Prinsloo & Lalor, 2020).

With respect to the role that delta synchronization plays in reading acquisition, the topographical distribution of the observed effect takes particular relevance. The temporal sampling framework (and its posterior modifications such as the amplitude modulation phase hierarchy perspective (Goswami, 2019)) assigns a role for delta in parsing and segmentation of the speech signal in auditory cortex. Thus, we would expect to observe a correlation between delta synchronization and reading at temporo-parietal electrodes. Our observation of this effect at frontal sites seems more in line with reflecting top-down modulations from frontal (including inferior frontal gyrus and precentral gyri) to temporal auditory regions. If this is so, then the difference between poor and typical readers observed here stems from the quality of the top-down linguistic or attentional modulation of auditory neural synchronization, and not through delta synchronization at the auditory level (Ding et al., 2015; Park et al., 2015). This would extend previous studies reporting differences in delta synchronization at the auditory level in older children and adults (Cutini et al., 2016; Hämäläinen et al., 2012; Molinaro et al., 2016; Power et al., 2016).

In sum, with respect to the predictions made by the temporal sampling framework, we did not find evidence for a theta-reading link, and, although we did find evidence for a delta-reading link, this did not seem to reflect differences in bottom-up temporal auditory processing but rather differences in top-down frontal modulations of auditory processing. To the best of our knowledge, this is the first study to show that delta synchronization in prereading children predicts future reading achievement one year later.

Chapter 7

Discussion

1. Summary of the findings

In the first study, we showed that it is both possible and feasible to identify children at risk of developing reading difficulties before the onset of reading acquisition itself. This contrasts with the current widespread practice of wait-for-failure approach that entails diagnosing dyslexia approximately in third grade. We know that remedial interventions are most effective the earlier they begin which demands timely identification of at-risk children. Screening can be done collectively, in the school setting, with minimal human and financial resources. This has profound consequences for setting in place remedial practices to reduce cascading effects of reading difficulties.

In the second study, we showed that reading acquisition in a transparent orthography such as Spanish follows a somewhat distinct trajectory than the one reported in English. In Spanish, the development of phonemic awareness skills is delayed but fast. Most children exhibit almost no phonemic awareness skills before reading acquisition, even in the presence of some letter knowledge, but most of them go on to achieve good decoding skills. The better their letter knowledge, verbal short-term memory and lexical access skills, the better readers they become. Nor phonemic neither syllabic awareness seems to contribute to explaining reading acquisition above and beyond the other preliteracy skills. This was not interpreted as PA having no role to play, but rather as a different, tighter association between PA and LK in Spanish.

In the third study, we showed that rhythmic sensitivity underlies reading acquisition in at least two ways. First, we found that, according to the available frameworks, rhythmic sensitivity underlies reading acquisition through its role in the formation of phonological representations. Second, and unexpectedly, we found that rhythmic sensitivity underlies reading acquisition independently from phonological processing. In trying to account for this effect, we —in line with other researchers— proposed that morphological awareness through lexical stress assignment might play a particularly important role in Spanish, where polysyllabic words are encountered early on during reading acquisition. Additionally, rhythmic sensitivity might reflect temporal processing skills, irrespective of its linguistic nature, in auditory processing. Finally, we found that rhythmic sensitivity improves identification of at-risk children, and that this gain is equivalent to the one obtained by phonological processing measures. Thus, rhythmic sensitivity emerges as a target for both screening and intervention.

In the fourth study, we showed that prereading children show neural synchronization to auditory non-speech stimuli at both delta and theta rates. Crucially, we showed that neural synchronization at delta rate in prereaders partially explain reading skills one year later, above and beyond Age and IQ. This finding provides novel evidence on the role that cortical oscillations play in auditory processing and reading acquisition. Interestingly, we observed the delta-reading link at frontal sites, compatible with top-down modulations of neural synchronization at the auditory cortex.

In the following section we integrate and discuss the main findings across studies.

2. Reading acquisition in a transparent orthography: emerging principles

2.1. On the contribution of phonological awareness

The reduced contribution of prereading phonological awareness to reading acquisition is an important, and controversial, finding in the present thesis, as discussed in Study Two. However, there is an apparent discrepancy on PA's contribution tor reading between Study One and Study Two. Several factors need to be considered.

First and foremost, sample composition in the two studies differed. While Study One included all children participating in the study, only children with no reading skills in K5 were included in Study Two. This was done specifically to address the issue of reciprocal effects between PA and reading (Castles & Coltheart, 2004; Huettig et al., 2018). The reasoning was the following: while children with reading difficulties often show accompanying poor PA skills, it is possible that these observed deficiencies in PA result from reduced or suboptimal reading experience. Thus, in order to test PA's unique contribution to reading acquisition, only children with no reading skills were studied, thus excluding the potential reciprocal effects of PA on reading. The discrepancy in significant predictors between Study One and Study Two in fact goes in line with this hypothesis: a quick additional exploration of Study's 1 data shows that, when the exact same prediction model is run *excluding all children with any reading skill in K5* (n = 54), the contribution of phonological awareness is reduced and the coefficient is no longer significant (estimate = - 0.76, SE = 0.41, statistic = -1.84, p = 0.065). Of course, it could be possible that the reduced coefficient reflects the decrease in sample size. However, given the theoretical arguments exposed above, the reciprocal PA-reading relationship stands as the most likely explanation.

Secondly, reading outcomes and modelling approach differed. With respect to reading outcome, in Study One we used a composite measure of decoding, fluency, and comprehension in order to predict overall reading performance. In contrast, in Study Two, two models were discussed, one for decoding and one for fluency, since the goal was to understand the cognitive underpinnings of reading acquisition, which might (and do) differ between decoding and fluency. While this might explain the observed discrepancies, it is not the most likely explanation, since it has been well documented that the contribution of PA to reading acquisition is stronger for decoding than for fluency or comprehension (Muter et al., 2004). If the discrepancies were related to the reading measure used as outcome, one would expect to find a strong phonological awareness contribution in the decoding model of Study Two, and less so for the composite measure of Study One. With respect to modelling approach, Study One used a logistic regression model in order to predict reading status (typical readers vs reading difficulties), while Study Two used a continuous linear regression model. These serve two different purposes. Linear regression is used to predict a continuous variable of outcomes (in this case, decoding or fluency), while logistic regression is used to classify outcomes into a binary category. It is expected that predicting variables would differ; it is not equivalent to discriminate between groups, than to predict an outcome variable in a continuous manner. For example, in Study One phonological awareness could predict class membership since, children who are already reading (K5 readers), might already show phonemic awareness skills, and nonreaders might not. Then, PA skills would explain class membership in Study One, but not decoding or fluency skills in Study Two, since K5 readers were not included in the sample.

Third, Study Two included School as a random effect, while Study One did not. Naturally, it would make no sense to include School in a screener since it would limit its generalizability capacities. In Study Two however, School is a most relevant variable since it accounts for the differences in reading scores that result from teaching practices (and not from individual differences). Thus, the explanatory effect of School could explain the different results between studies.

In sum, differences in sample composition, modelling approach and inclusion of the school variable are likely to explain the discrepancy between significant predictors in Study One and Study One. Overall, we agree with Meteyard and Davies (2020) approach to model building in that researchers should "acknowledge that the choices you make during analysis are considered, justified and one path amongst many" (Meteyard & Davies, 2020, p. 20). Taken together, the reported evidence suggests that the contribution of prereading phonological awareness to reading acquisition in a transparent orthography is reduced with respect to that reported for English, and that this could be explained by its delayed development (Figure 7.1) and also by a tighter association with letter knowledge skills.

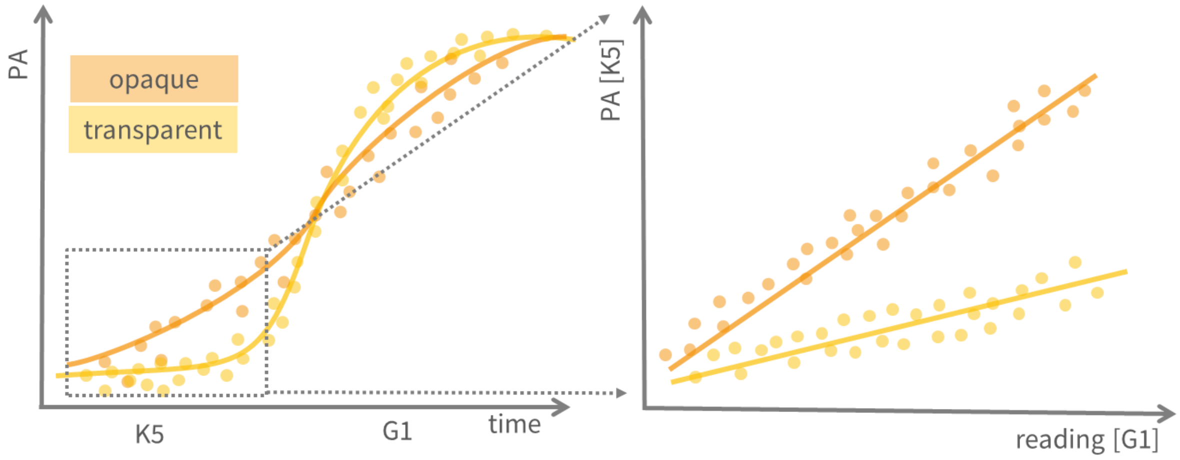


Figure 7.1. Schematic illustration of phonological awareness and reading acquisition across orthographies. Left: Development of phonological awareness skills in an opaque and a transparent orthography. In a transparent orthography, we propose that phonological awareness develops late but fast. Right: Prereading phonological awareness relation to first grade reading. In transparent orthographies, we propose that the correlation is reduced due to the floor effects of prereading phonological awareness. If early K5 readers were included in the sample, the distribution of prereading phonological awareness scores would become wider, and thus the correlation would increase (not depicted).

2.2. On the contribution of verbal short-term memory

Along the presented studies, verbal short-term memory has popped in and out playing a relevant role during reading acquisition. Study One showed no contribution of verbal short-term memory in classifying children into poor and typical readers groups using a composite measure of reading which included decoding, fluency, and comprehension. Study Two showed a relevant contribution for vSTM on predicting continuous reading outcomes for decoding but not for fluency. Study Three showed that vSTM mediates the rhythm-reading relationship when reading is operationalized as efficiency, a ratio between decoding accuracy and response time. There seems to be an emerging picture in which verbal short-term memory plays a more important role in decoding than in fluency or comprehension. These findings go in line with conceptualizations of decoding and fluency. Decoding entails applying grapheme to phoneme conversion rules to each individual letter, maintaining them in memory, and then blending phonemes together. Such a process implies a considerable implication of verbal short-term memory. Fluency, on the other hand, is more dependent on the automaticity of visual word recognition and visuo-attentional skills for targeting word sequences, a process much better indexed by RAN (Altani et al., 2020).

An additional aspect of verbal short-term memory needs to be considered. A large early modelling analysis by Wagner and colleagues showed that while phonological awareness and phonological coding in working memory (equivalent to what we have termed verbal short-term memory) were better represented as a single construct during kindergarten, by second grade they were better represented as two distinct constructs (Torgesen et al., 1994). Additionally, when modelling these two variables —measured in kindergarten— as predictors of grade 1 reading achievement along with RAN measures, they found that the only one who contributed unique variance was phonological awareness. They end their report with a crucial methodological consideration, that is too often forgotten:

> When variables that are correlated with one another (as were all the phonological variables) are included in simultaneous causal equation, a predictor that is only slightly more strongly related to the criterion can receive a substantial coefficient, while a second, correlated predictor receives a coefficient near zero because it does not make a causal contribution that is unique from the first predictor. This does not mean that the second variables is not causally related to the criterion; it simply means that the causal contribution of the two variables are redundant (Torgesen et al., 1994, p. 284)

Lastly, a second look at the verbal short-term memory task can shed light on its involvement in reading acquisition. The most frequently used task in assessing verbal-short term memory (sometimes called phonological short-term memory) is pseudoword repetition. In this task participants are presented with a sequence of pseudowords increasing in length and are asked to repeat them back. This entails recall of both item and order information. In our study, we tried to avoid verbal responses in order to have a screening battery with minimal involvement of trained personal, and that could be self-administered. Thus, we employed a different task based on Martinez-Perez et al.'s work (Martinez Perez et al., 2012). In this task, children would hear a sequence of monosyllabic words followed by their corresponding images. Their task was to order the images in the same order that they heard the words. For example, they heard /sol/, /pan/, /pez/, saw three images depicting a sun, a bread, and a fish (in random order), and had to arrange them on screen from left to right. The sequence would increase in length from two to seven items. Since items were visually displayed on screen after hearing them, there was no item-recall memory involved. Instead, they only needed to recall the serial order information. In this sense, our task relied less on recoding phonological information than typical short-term memory tasks. In a longitudinal study with prereaders, Martinez-Perez et al. showed that verbal short-term memory for order contributed unique variance to decoding above and beyond verbal short-term memory for items and also above and beyond phonological awareness (Martinez Perez et al., 2012). However, in their study verbal short-term memory for items did not contribute unique variance above phonological awareness. They concluded that while verbal short-term memory for items is closely related to phonological representations and processes, verbal short-term memory for order reflects additional processing mechanisms, such as temporarily maintaining a sequence representation. This could explain why we observe a unique and significant contribution of verbal short-term memory to decoding, above the variance explained by phonological awareness.

2.3. On the contribution of rhythmic sensitivity

Studies Three and Four both addressed the link between rhythmic sensitivity and reading from two complementary points of view, at the cognitive neural levels. In Study Three, we showed that rhythmic sensitivity, behaviourally measured, subserved reading acquisition. We interpreted this in terms of phonological processing (phonological awareness and verbal-short term memory) and also beyond it, through stress assignment. These may represent two components of stress sensitivity at the segmental and suprasegmental (i.e., prosodic) levels, respectively. Thus, the overall emerging picture suggest that the role of phonology in the rhythm-reading relationship should be conceptualized in its broader sense, including both segmental and suprasegmental components. Findings from Study Four confirmed the rhythm-reading relationship from a neural perspective. It showed that rhythmic sensitivity at the delta rate, corresponding to supralexical-prosodic information, explains reading acquisition one year later.

Although not directly isolated in our paradigms, *temporal processing* underlies phonological skills in its broader sense. Both the tapping to a beat task, and brain-speech synchronization to rhythmic auditory stimuli have been shown to reflect temporal processing skills (Arnal et al., 2014; Ozernov-Palchik et al., 2018). This is particularly relevant for our results, since the stimuli used in both studies were non-linguistic. These results point to the putative relevance of temporal processing as an underlying mechanism in reading acquisition.

In addition to temporal processing, *short-term memory* also emerges as a candidate mediator in the rhythm-reading relation. On one hand, finding from Study Three show that when verbal-short term memory is included in the rhythm-reading model, the contribution of rhythm is substantially (although not completely) reduced. On the other hand, regarding Study Four, chunking (arguably, a form of memory) has been shown to underly brain-speech synchronization at the delta rate (Boucher et al., 2017; Meyer, 2018). Thus, results from both studies point to the role that memory might play in the rhythm-reading relation.

Taken together, findings from studies 3 and 4 underscore the role that rhythmic sensitivity plays in reading acquisition, and suggest possible mediators, namely phonological awareness, verbal short-term memory, lexical stress assignment, and temporal processing (Figure 7.2). Of course, these constructs are tightly related to each other, and thus together they point to a possible pathway from rhythm to reading. Crucially, since the reported effects stem from a longitudinal design with prereaders, both studies add evidence in line with a *causal* relationship between rhythmic sensitivity and reading.

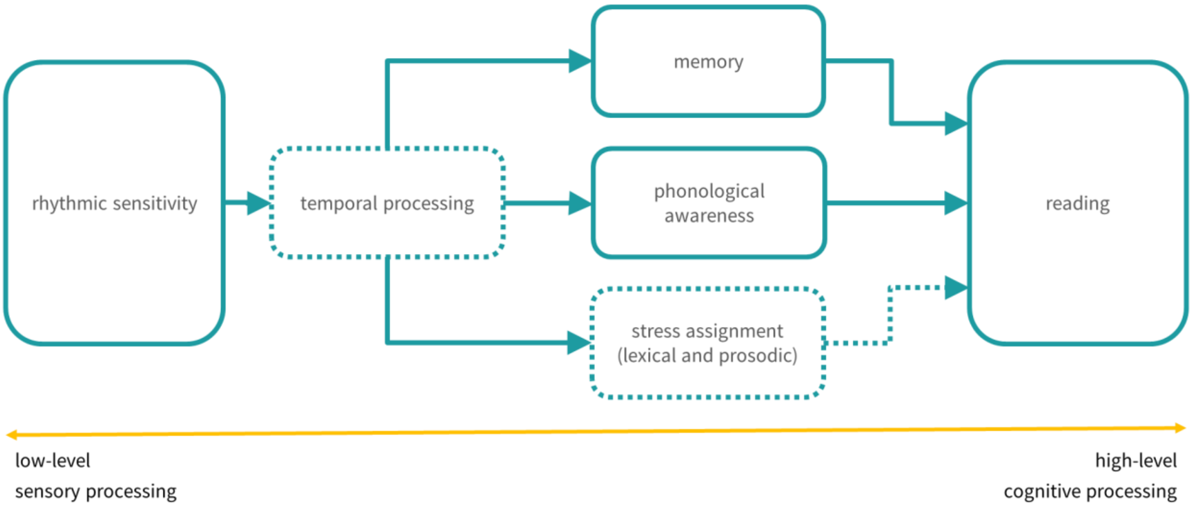


Figure 7.2. Schematic illustration of mediators between rhythmic sensitivity and reading acquisition. Continuous lines represent links tested in the present thesis; dashed lines represent candidate mediators based on previous literature. The paths are of course not independent and also not exhaustive.

3. Outstanding questions and future directions

A necessary follow-up from Study One is to validate the *reduced* Lexiland battery, and to test it in a broader sample including Schools from low SES. In parallel, it is crucial to start with its dissemination and agreements with the Public School System for it to be implemented nationwide.

Findings from Study Two leave the open question of exactly how letter knowledge and phonological awareness are reciprocally modified, especially in the 4- to 6-year-old period. While this issue has been thoroughly studied in English (e.g., Treiman & Kessler, 2004), much less evidence is available for Spanish. Given the vast differences in grapheme-to-phoneme correspondences across orthographies, it is expected that the LK-PA relation will develop differently across orthographies. It is possible that in transparent orthographies like Spanish, if a child has enough letter knowledge and memory, learning to blend sounds (i.e., PA) is equivalent to reading, whereas in opaque orthographies this is an intermediate step. We plan to use computational models (e.g., Coltheart et al., 2001; Harm & Seidenberg, 2004) to understand whether this explains the different results reported across orthographies. If our interpretation is correct (Figure 7.1), we should expect to see rapid changes in phonological awareness in the transitions to learning to read. A study following children with more frequent evaluations could be used to reveal the sudden transition. Moreover, evaluating the response to specific interventions could also provide valuable information about the causal relations involved.

Important open questions remain from Study Three and Study Four. Namely, what underlying processes does rhythmic sensitivity represent? The first next step is to examine the relationship between rhythmic sensitivity behaviourally and neurally measured, an enterprise that can be easily done with the available data. An additional open question is what the mediators between rhythmic sensitivity and reading are. We have shown that phonological processing explains some of its contribution to reading, but we have also shown that this is not the whole picture. Building on previous studies, we have suggested that prosodic and lexical stress assignment at the cognitive/linguistic level, as well as temporal processing could underly the rhythm-reading relation. These would need further testing. To begin with, it would be possible with the available data to study the association between rhythmic sensitivity neurally measured, and phonological awareness and verbal short-term memory. A different strand of questions stem from how linguistic and non-linguistic rhythmic skills relate to each other, and how each relates to reading. We have found important contributions of non-linguistic rhythmic sensitivity to reading acquisition, does linguistic rhythmic sensitivity explain additional variance to reading acquisition? Previous studies suggest it would (Ozernov-Palchik et al., 2018). Finally, we have completely left aside the role that asymmetric hemispheric processing of auditory stimuli plays in phonological development and reading acquisition (Boemio et al., 2005). It would be possible to test this hypothesis with the available EEG data, there is currently very scarce evidence coming from developmental designs (e.g., Rios Lopez, 2018).

4. Conclusions

The discovery of the importance of phonological representations to reading acquisition produced a revolution in our understanding of reading, reading instruction, and dyslexia (Bradley & Bryant, 1983). Most of the studies supporting this view came from non-transparent orthographies. In line with the movement towards the expansion of Cognitive Science and Neuroscience to embrace the diversity of populations (Henrich et al., 2010), this thesis provides relevant evidence that the broad theoretical picture has a particular instantiation when applied to a transparent orthography. The results presented do not question the relevance of phonology for reading acquisition. In fact, the participation of brain oscillations and their manifestations in the prediction of reading points to the fact that on top of a common neurobiology, the particular details can vary when they have to accommodate the specifics of the cultural environment. Oscillatory processes allow parsing the sound signal in precise phonological representations that are crucial for the learning of letter-sound correspondences, and possibly for memory processes. The same processes support the development of phonological awareness, required for reading acquisition in English, and almost equivalent to decoding in Spanish. We provide novel infrequent evidence —from a longitudinal perspective and a transparent orthography— of how these processes develop and interact. We believe this adds to the broader picture of the universality of reading acquisition, and its instantiation in different orthographies.

Resumen extendido en castellano

La presente tesis estudia los sustratos cognitivos y neurales del aprendizaje de la lectura en una ortografía transparente. En cuatro estudios, aborda el problema del aprendizaje de la lectura desde una perspectiva traslacional combinando el estudio del monitoreo del riesgo lector en el contexto escolar, el estudio de las características distintivas de las bases cognitivas del aprendizaje de la lectura en una ortografía transparente, y el estudio de la sincronización neural a estímulos auditivos rítmicos en el laboratorio.

Estos objetivos se lograron a través de un diseño longitudinal comenzando desde la educación infantil, siguiendo a un mismo conjunto de niños hasta segundo año de escuela. Los niños fueron evaluados en grupo, en el contexto escolar, a través de una App —*Lexiland*— desarrollada en el marco de la presente tesis. Esta App nos permitió evaluar en un tiempo corto a una gran muestra de aproximadamente 600 niños. La evaluación se centró en habilidades de alfabetización emergente (conciencia fonológica, conocimiento de las letras, velocidad de denominación) y habilidades cognitivas generales (memoria de corto plazo, cociente intelectual y vocabulario). Se evaluó también la sensibilidad al ritmo, por su presunto rol en el desarrollo fonológico, y se obtuvieron además medidas demográficas, a destacar, el nivel socioeconómico de los niños participantes.

A partir de estos datos, y de medidas neurales tomadas en una submuestra de los niños, se abordaron cuatro preguntas principales, que desarrollamos a continuación.

1. El primer problema: prácticas actuales en la detección de dificultades lectoras

Las dificultades en la lectura tienen consecuencias en cascada para el desarrollo socioemocional y profesional de quienes las padecen (Arnold et al., 2005). Las prácticas actuales en la detección de dificultades en la lectura generalmente implican esperar hasta el tercer grado de la escuela primaria para alcanzar un diagnóstico de dislexia y ofrecer intervenciones adecuadas (Ozernov-Palchik & Gaab, 2016). Por muy conservadora que sea esta estrategia, que es una salvaguardia para evitar el sobrediagnóstico y los falsos positivos, sus consecuencias prácticas son demasiado perjudiciales para sostenerla. Desde hace décadas, ha habido suficiente conocimiento para predecir, con más o menos sensibilidad, las trayectorias futuras de lectura a partir de la evaluación de algunas habilidades críticas que se desarrollan antes de la instrucción formal de lectura (ver, por ejemplo, Lyytinen et al., 2015). Sin embargo, este conocimiento no ha dado lugar a estrategias de prevención adecuadas. Asimismo, actualmente la evidencia muestra que las intervenciones educativas son más efectivas cuanto antes comienzan (Wanzek & Vaughn, 2007). Por lo tanto, el enfoque actual de esperar al fracaso tiene graves consecuencias a mediano y largo plazo.

El primer estudio de la presente tesis busca propone una alternativa el abordaje actual a través de desarrollar un instrumento de tamizaje que permite identificar a niños en riesgo lector cuando están aún cursando educación infantil. Este instrumento se desarrolló como un videojuego —*Lexiland*— que evalúa habilidades de alfabetización emergente. Es digital, autoadministrada, breve y puede utilizarse en el contexto de aula. Se evaluó su poder predictor a través del seguimiento longitudinal de una muestra de aproximadamente 600 niños desde educación infantil hasta segundo año de educación primaria. Durante educación infantil se evaluó el conocimiento de las letras, la conciencia fonológica, la velocidad de denominación, la memoria verbal y no verbal, el vocabulario y el cociente intelectual. En primero y segundo año se evaluó la decodificación de palabras y pseudopalabras, y la fluidez y comprensión lectoras.

Los resultados muestran que *Lexiland* es capaz de predecir las dificultades lectoras futuras con altos niveles de sensibilidad y especificidad. Más aún, solo cuatro variables son suficientes para alcanzar altos niveles de precisión en la clasificación, lo cual deriva en un tamizaje breve y plausible de ser utilizado en el contexto de aula. Además, dado que la evaluación es digital, grupal y autoadministrada, su bajo costo hace que sea plausible de ser utilizada a nivel nacional. Por lo tanto, los hallazgos demuestran que la detección oportuna del riesgo lector puede y debe realizarse durante la educación infantil si se desean prevenir algunas de las trayectorias de riesgo lector.

2. El segundo problema: controversias actuales sobre la universalidad de los procesos cognitivos que subyacen al aprendizaje de la lectura

La universalidad de los sustratos cognitivos del aprendizaje de la lectura es actualmente un tema de controversia. La mayoría de la evidencia proviene de estudios llevados a cabo con personas de habla inglesa que aprenden a leer en inglés (Share, 2008). El inglés posee una ortografía que puede considerarse como opaca, ya que las reglas de correspondencia entre grafemas y fonemas son muy inconsistentes. El español, en cambio, posee una ortografía transparente, esto es, las reglas de correspondencia entre grafemas y fonemas son muy consistentes. Dado el importante rol que juega la conversión de grafemas a fonemas en las etapas iniciales del aprendizaje de la lectura, cabe esperarse que la consistencia de la ortografía sea un factor modulador del proceso de aprendizaje de la lectura, y así lo prueban múltiples estudios (ver, por ejemplo, Seymour et al., 2003). En el segundo estudio investigamos hasta qué punto pueden traducirse los hallazgos de ortografías inconsistentes a una ortografía consistente como la del español. No solo por la necesidad de producir conocimiento local para abordar los problemas locales, sino también para enriquecer la discusión académica a través de evidencias novedosas y poco frecuentes.

Los datos obtenidos en el primer estudio se utilizaron para construir un modelo longitudinal de predicción del desempeño lector, con una gran muestra y un control exhaustivo de posibles variables de confusión (*confound variables*). Se construyó un modelo para predecir el desempeño en decodificación y otro para predecir el desempeño en fluidez, y se excluyeron de la muestra aquellos niños que durante educación infantil ya sabían leer, de modo de excluir relaciones recíprocas entre las habilidades de alfabetización emergentes y la lectura (Huettig et al., 2018).

Los resultados muestran que el papel de las habilidades de alfabetización emergente (conciencia fonológica, conocimiento de letras, velocidad de denominación) difiere en español en comparación con otras ortografías menos consistentes. En particular, la conciencia fonológica parece contribuir a la predicción del desempeño lector futuro más allá de otras habilidades de alfabetización emergente. A su vez, el conocimiento de las letras asume un papel más central en la predicción del desempeño lector futuro. Proponemos que estos hallazgos pueden explicarse a través de una trayectoria de desarrollo más lenta para la conciencia fonológica (modulada por las actividades de alfabetización en el hogar y las prácticas educativas), las características intrínsecas del sistema ortográfico, un papel importante para la memoria verbal a corto plazo, y una asociación más estrecha entre el conocimiento de las letras y la conciencia fonológica.

3. El tercer problema: bases cognitivas del procesamiento fonológico

El procesamiento fonológico ha estado en el centro de la comprensión mecanicista de las dificultades de lectura (Melby-Lervåg et al., 2012). Si bien un puñado de teorías explican las dificultades de lectura a través de mecanismos adicionales (visuoespacial, atencional, motor y en el aprendizaje o anclaje estadístico), el papel de la fonología es indiscutible. Cabe preguntarse entonces cuál es la base sensoriocognitiva del procesamiento fonológico. Un buen candidato desde los primeros días de los estudios de procesamiento fonológico es el ritmo (Cutler & Mehler, 1993). La sensibilidad rítmica es la base del procesamiento del acento en el habla, que a su vez juega un papel central en la segmentación del habla, que es la base de la formación de representaciones fonológicas. Por lo tanto, es de esperar que el procesamiento del ritmo impacte en el futuro aprendizaje de la lectura a través de su rol en el procesamiento fonológico. Si bien algunos estudios muestran evidencias de las asociación entre la sensibilidad rítmica y el desempeño lector (Leong & Goswami, 2014), la evidencia proveniente de niños pre-lectores es escasa.

En el tercer estudio investigamos la asociación entre la sensibilidad rítmica, el procesamiento fonológico y el futuro desempeño lector. Para ello evaluamos durante educación infantil la capacidad de los niños de tamborilear sincrónicamente a un estímulo rítmico auditivo (*beat*) a tres frecuencias diferentes (60 bpm, 120 bpm y 240 bpm), y relacionamos su desempeño con sus habilidades fonológicas (conciencia fonológica y memoria verbal de corto plazo) y su futuro desempeño lector en decodificación.

Los resultados muestran que la sensibilidad rítmica predice el futuro desempeño lector a través del procesamiento fonológico, pero también independientemente de él. Sugerimos que este último efecto puede reflejar el papel de la sensibilidad rítmica en el procesamiento fonológico a niveles suprasegmentales, tales como la prosodia, y, por lo tanto, que la relación ritmo-lectura debe ser entendida en un marco más amplio que el de la conciencia fonológica. En segundo lugar, encontramos que la sensibilidad rítmica permite identificar a niños en riesgo lector cuando es evaluada usando *Lexiland* en el contexto de aula, lo cual la posiciona como un sólido candidato en el tamizaje para la identificación de dificultades en la lectura.

4. El cuarto problema: fundamentos neurales del procesamiento fonológico y sus consecuencias sobre el desempeño lector

La sensibilidad rítmica y las representaciones fonológicas dependen, en general, del procesamiento auditivo, y, en particular, del procesamiento auditivo a modulaciones temporales o ritmos. Tanto modelos computacionales como evidencias empíricas muestran que las oscilaciones cerebrales en la corteza auditiva se sincronizan con los ritmos del habla (especialmente en 2, 4 y aproximadamente 30 Hz), y que este mecanismo subyace a la segmentación del habla y por lo tanto al desarrollo de representaciones fonológicas (Giraud & Poeppel, 2012; Gross et al., 2013). Si bien muchos estudios han abordado este problema en adultos y niños mayores, apenas unos pocos lo han estudiado en pre-lectores (Lizarazu et al., 2021). Dado que el aprendizaje de la lectura tiene efectos recíprocos sobre las representaciones fonológicas, cabe preguntarse si existe una relación entre la sincronización de las oscilaciones cerebrales a estímulos auditivos y el desempeño lector, incluso antes del aprendizaje de la lectura, en niños pre-lectores. Por tanto, en el cuarto estudio examinamos el procesamiento neural de estímulos rítmicos auditivos y su relación con el futuro desempeño lector.

Con ese objetivo, registramos la actividad neural de aproximadamente 40 niños pre-lectores a través de un estudio electroencefalográfico en el que los niños escuchaban ruido blanco modulado en amplitud a tres frecuencias diferentes (2, 4 y 8 Hz). Estudiamos la sincronización entre las oscilaciones cerebrales y los estímulos auditivos, y, posteriormente, la relación entre esta sincronización —medida durante educación infantil— y el desempeño lector medido un año después, en primer año de primaria.

Los resultados muestran, en primer lugar, que en niños pre-lectores se observa una sincronización entre oscilaciones cerebrales y estímulos auditivos a 2 y 4 Hz, pero no a 8 Hz. Esto pone de manifiesto la especificidad de la sincronización neural en las frecuencias relevantes para el procesamiento del habla (2 y 4 Hz), pero no para frecuencias irrelevantes. En segundo lugar, observamos que la sincronización neural a 2 Hz, pero no a 4 Hz, se asocia al futuro desempeño lector. Estos hallazgos proporcionan evidencia novedosa sobre el papel que juegan las oscilaciones cerebrales en el procesamiento auditivo y el aprendizaje de la lectura.

Conclusiones

A lo largo de los cuatro estudios se abordó el problema del aprendizaje de la lectura, enlazando niveles de análisis desde el tamizaje de riesgo lector en contexto escolar hasta las bases neurales del procesamiento auditivo y su relación con el futuro desempeño lector en contexto de laboratorio.

Una de las principales contribuciones de los estudios proviene del diseño longitudinal, que proporciona evidencia a favor de un rol causal de los sustratos neurocognitivos que subyacen a la adquisición de la lectura. En segundo lugar, ofrece evidencia sobre el aprendizaje de la lectura en una ortografía transparente, la cual es poco frecuente. En tercer lugar, aborda estos problemas en un contexto ecológico al evaluar a los niños en el entorno escolar, a través de un videojuego desarrollada para este mismo fin. Tal enfoque hizo posible evaluar una gran muestra de aproximadamente 600 niños de educación infantil, aumentando la potencia del estudio.

En resumen, presentamos un estudio longitudinal sobre los sustratos cognitivos y neuronales del aprendizaje de la lectura en una ortografía transparente, combinando estudios en contexto de aula y en contexto de laboratorio.

References

Abusamra, V., Cartoceti, R., Raiter, A., & Ferreres, A. (2008). Una perspectiva cognitiva en el estudio de la comprensión de textos. *Psico*, *39*(3), 352–361.

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology*, *112*(1), 22–39. https://doi.org/10.1037/edu0000359

Andrade, O. V. C. A., Andrade, P. E., & Capellini, S. A. (2015). Collective screening tools for early identification of dyslexia. *Frontiers in Psychology*, *5*(January), 1–13. https://doi.org/10.3389/fpsyg.2014.01581

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly*, *38*(4), 470–487.

Anthony, J. L., Williams, J. M., Durán, L. K., Gillam, S. L., Liang, L., Aghara, R., Swank, P. R., Assel, M. A., & Landry, S. H. (2011). Spanish Phonological Awareness: Dimensionality and Sequence of Development During the Preschool and Kindergarten Years. *Journal of Educational Psychology*, *103*(4), 857–876. https://doi.org/10.1037/a0025024

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, *83*(2), 111–130. https://doi.org/10.1016/S0022-0965(02)00124-8

Araújo, S., Fernandes, T., & Huettig, F. (2019). *Learning to read facilitates the retrieval of phonological representations in rapid automatized naming : Evidence from unschooled illiterate , ex - illiterate , and schooled literate adults*. *July 2018*, 1–9. https://doi.org/10.1111/desc.12783

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology*, *107*(3), 868–883. https://doi.org/10.1037/edu0000006

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2014). Delta-Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. *Cerebral Cortex*, *2*, 1–9. https://doi.org/10.1093/cercor/bhu103

Arnold, E. M., Goldston, D. B., Walsh, A. K., Reboussin, B. A., Daniel, S. S., Hickman, E., & Wood, F. B. (2005). Severity of Emotional and Behavioral Problems Among Poor and Typical Readers. In *Journal of Abnormal Child Psychology* (Vol. 33, Issue 2, pp. 205–217). https://doi.org/10.1007/s10802-005-1828-9

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Goswami, U., Grey, C., & Flanagan, S. (2020). Delta- And theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *BioRxiv*, 1–36. https://doi.org/10.1101/2020.10.12.329326

Baker, F. B. (2001). *The Basics of Item Response Theory* (C. Boston & L. Rudner (eds.); Second). ERIC Clearninghouse on Assessment and Evaluation.

Baroni, F., Morillon, B., Trébuchon, A., Liégeois-Chauvel, C., Olasagasti, I., & Giraud, A.-L. (2019). Converging intracortical signatures of two separated processing timescales in human early auditory cortex. *BioRxiv*, 1–35. https://doi.org/10.1101/730002

Barron, R. W. (1991). Proto-literacy, literacy and the acquisition of phonological awareness. *Learning and Individual Differences*, *3*(3), 243–255. https://doi.org/10.1016/1041-6080(91)90010-X

Bartoń, K. (2019). MuMIn: Multi-Model Inference, Version 1.43.6. *R Package*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*(3), 389–395. https://doi.org/10.1038/nn1409

Boets, B., Wouters, J., van Wieringen, A., & Ghesquière, P. (2006). Auditory temporal information processing in preschool children at family risk for dyslexia: relations with phonological abilities and developing literacy skills. *Brain and Language*, *97*(1), 64–79. https://doi.org/10.1016/j.bandl.2005.07.026

Boets, B., Wouters, J., van Wieringen, A., & Ghesquière, P. (2007). Auditory processing, speech perception and phonological ability in pre-school children at high-risk for dyslexia: a longitudinal study of the auditory temporal processing theory. *Neuropsychologia*, *45*(8), 1608–1620. https://doi.org/10.1016/j.neuropsychologia.2007.01.009

Boles, D. B., & Clifford, J. E. (1989). An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, Instruments, & Computers*, *21*(6), 579–586. https://doi.org/10.3758/BF03210580

Boucher, V., Gilbert, A., & Jemel, B. (2017). The Role of Low-frequency Neural Oscillations in Speech Processing: Revisiting Delta Entrainment. *Journal of Cognitive Neuroscience*, 139. https://doi.org/10.1162/jocn

Bowers, P. G., & Wolf, M. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*(3), 415–438. https://doi.org/10.1037/0022-0663.91.3.415

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., Götz, K., & Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*(4), 422–432. https://doi.org/10.1111/j.1469-7610.2007.01849.x

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read - A causal connection. *Nature*, *301*(5899), 419–421. https://doi.org/10.1038/301419a0

Cabell, S. Q., Justice, L. M., Zucker, T. A., & Kilday, C. R. (2009). Validity of teacher report for assessing the emergent literacy skills of at-risk preschoolers. *Language, Speech, and Hearing Services in Schools*, *40*(2), 161–173. https://doi.org/10.1044/0161-1461(2009/07-0099)

Calet, N., Gutiérrez-Palma, N., Simpson, I. C., González-Trujillo, M. C., & Defior, S. (2015). Suprasegmental Phonology Development and Reading Acquisition: A Longitudinal Study. *Scientific Studies of Reading*, *19*(1), 51–71. https://doi.org/10.1080/10888438.2014.976342

Calonge, N. (2008). Universal screening for hearing loss in newborns: US preventive services task force recommendation statement. *Pediatrics*, *122*(1), 143–148. https://doi.org/10.1542/peds.2007-2210

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A Cross-Linguistic, Longitudinal Study of the Foundations of Decoding and Reading Comprehension Ability. *Scientific Studies of Reading*, *23*(5), 386–402. https://doi.org/10.1080/10888438.2019.1580284

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavsky, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science*, *23*(6), 678–686. https://doi.org/10.1177/0956797611434536

Cardoso-Martins, C., Mesquita, T. C. L., & Ehri, L. (2011). Letter names and phonological awareness help children to learn letter-sound relations. *Journal of Experimental Child Psychology*, *109*(1), 25–38. https://doi.org/10.1016/j.jecp.2010.12.006

Cardozo, S. (2016). *Trayectorias educativas en la educación media PISA-L 2009-2014*. INEEd - Grupo de estudios sobre Transiciones Educación-Trabajo (TET).

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77–111. https://doi.org/10.1016/S0010-0277(03)00164-1

Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, *42*(2), 163–176. https://doi.org/10.1177/0022219408326219

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*(2), 300–315. https://doi.org/10.1037/a0033805

Colling, L. J., Noble, H. L., & Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An EEG study. *Frontiers in Neuroscience*, *11*(JUL). https://doi.org/10.3389/fnins.2017.00360

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. http://www.ncbi.nlm.nih.gov/pubmed/11212628

Corsi, P. (1972). Memory and the medial temporal region of the brain. *Unpublished Doctoral Dissertation), McGill University, Montreal, QB*.

Cuetos, F., Rodríguez, B., Ruano, E., & Arribas, D. (2007). Prolec-R, Batería de evaluación de los procesos lectores, Revisada. *Madrid: TEA*.

Cunningham, A. J., Witton, C., Talcott, J. B., Burgess, A. P., & Shapiro, L. R. (2015). Deconstructing phonological tasks: The contribution of stimulus and response type to the prediction of early decoding skills. *Cognition*, *143*, 178–186. https://doi.org/10.1016/j.cognition.2015.06.013

Cutini, S., Szűcs, D., Mead, N., Huss, M., & Goswami, U. (2016). Atypical right hemisphere response to slow temporal modulations in children with developmental dyslexia. *NeuroImage*, *143*, 40–49. https://doi.org/10.1016/j.neuroimage.2016.08.012

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, *21*(1–2), 103–108. https://doi.org/10.1016/s0095-4470(19)31323-3

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, *25*(4), 385–400. https://doi.org/10.1016/0749-596X(86)90033-1

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*(3), 381–410. https://doi.org/10.1016/0010-0285(92)90012-Q

De Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST Dataset of Language Proficiency. *Frontiers in Psychology*, *8*(April). https://doi.org/10.3389/fpsyg.2017.00522

De Jong, P. F., & Van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in Dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*(1), 22–40. https://doi.org/10.1037/0022-0663.95.1.22

De Vos, A., Vanvooren, S., Ghesquière, P., & Wouters, J. (2020). Subcortical auditory neural synchronization is deficient in pre-reading children who develop dyslexia. *Developmental Science*, *23*(6), 1–16.

https://doi.org/10.1111/desc.12945

De Vos, A., Vanvooren, S., Vanderauwera, J., Ghesquière, P., & Wouters, J. (2017). A longitudinal study investigating neural processing of speech envelope modulation rates in children with (a family risk for) dyslexia. *Cortex*, *93*, 206–219. https://doi.org/10.1016/j.cortex.2017.05.007

Defior, S., Serrano, F., & Marín-Cano, M. J. (2008). El poder predictivo de las habilidades de conciencia fonológica en la lectura y escritura en castellano. En. *Estudios de Desarrollo Del Lenguaje y Educación. Studies on Language Development and Education*, *January*, 339–347.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. In *Cognition* (Vol. 77, Issue 3). https://doi.org/10.1016/S0010-0277(00)00106-2

Dresler, T., Bugden, S., Gouet, C., Lallier, M., Oliveira, D. G., Pinheiro-Chagas, P., Pires, A. C., Wang, Y., Zugarramurdi, C., & Weissheimer, J. (2018). A Translational Framework of Educational Neuroscience in Learning Disorders. *Frontiers in Integrative Neuroscience*, *12*(July), 1–12. https://doi.org/10.3389/fnint.2018.00025

Duncan, L. G., Castro, S. L., Defior, S., Seymour, P. H. K., Baillie, S., Leybaert, J., Mousty, P., Genard, N., Sarris, M., Porpodas, C. D., Lund, R., Sigurdsson, B., Thráinsdóttir, A. S., Sucena, A., & Serrano, F. (2013). Phonological development in relation to native language and literacy: Variations on a theme in six alphabetic orthographies. *Cognition*, *127*(3), 398–419. https://doi.org/10.1016/j.cognition.2013.02.009

Dunn, L. M., Dunn, L. M., & Arribas, D. (2006). Peabody, test de vocabulario en imágenes. *Madrid: Tea*.

Ehri, L. C. (2008). Development of Sight Word Reading: Phases and Findings. In *The Science of Reading: A Handbook* (pp. 135–154). John Wiley and Sons. https://doi.org/10.1002/9780470757642.ch8

Elbro, C. (1996). Early linguistic abilities and reading development: A review and a hypothesis. In *Reading and Writing* (Vol. 8, Issue 6, pp. 453–485). https://doi.org/10.1007/BF00577023

Etard, O., & Reichenbach, T. (2019). Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *39*(29), 5750–5759. https://doi.org/10.1523/JNEUROSCI.1828-18.2019

Farahani, E. D., Goossens, T., Wouters, J., & van Wieringen, A. (2017). Spatiotemporal reconstruction of auditory steady-state responses to acoustic amplitude modulations: Potential sources beyond the auditory pathway. *NeuroImage*, *148*, 240–253. https://doi.org/10.1016/j.neuroimage.2017.01.032

Farahani, E. D., Wouters, J., & Wieringen, A. Van. (2020). Brain mapping of auditory steady-state responses : A broad view of cortical and subcortical sources. *Human Brain Mapping*, *October*, 1–17. https://doi.org/10.1002/hbm.25262

Flo, E. (2019). *Elementary composition in spanish : an EEG study*. Universidad de la República.

Foster, J., Barkus, E., & Yavorsky, C. (2011). Understanding and Using Advanced Statistics. In *Understanding and Using Advanced Statistics* (Issue August). https://doi.org/10.4135/9780857020154

Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing*, *18*(2), 129–155. https://doi.org/10.1007/s11145-004-5892-2

Furnes, B., Elwér, Å., Samuelsson, S., Olson, R. K., & Byrne, B. (2019). Investigating the Double-Deficit Hypothesis in More and Less Transparent Orthographies: A Longitudinal Study from Preschool to Grade 2. *Scientific*

Studies of Reading, *23*(6), 478–493. https://doi.org/10.1080/10888438.2019.1610410

Furnes, B., & Samuelsson, S. (2010). Predicting reading and spelling difficulties in transparent and opaque orthographies: *Dyslexia*, *239*, 234–239. https://doi.org/10.1002/dys

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. https://doi.org/10.1007/s11145-010-9271-x

Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience*, *32*(5), 545–561. https://doi.org/10.1080/23273798.2016.1232419

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*(1–2), 113–126. https://doi.org/10.1159/000208934

Giraud, A L, & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Giraud, Anne Lise, Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, *84*(3), 1588–1598. https://doi.org/10.1152/jn.2000.84.3.1588

González-Trujillo, M. C., Defior, S., & Gutiérrez-Palma, N. (2014). The role of nonspeech rhythm in Spanish word reading. *Journal of Research in Reading*, *37*(3), 316–330. https://doi.org/10.1111/j.1467-9817.2012.01529.x

Goodman, I., Libenson, A., & Wade-Woolley, L. (2010). Sensitivity to linguistic stress, phonological awareness and early reading ability in preschoolers. *Journal of Research in Reading*, *33*(2), 113–127. https://doi.org/10.1111/j.1467-9817.2009.01423.x

Goswami, U. (2003). Why theories about developmental dyslexia require developmental designs. *Trends in Cognitive Sciences*, *7*(12), 534–540. https://doi.org/10.1016/j.tics.2003.10.003

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. https://doi.org/10.1016/j.tics.2010.10.001

Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, *16*(1), 43–54. https://doi.org/10.1038/nrn3836

Goswami, U. (2018). A Neural Basis for Phonological Awareness? An Oscillatory Temporal-Sampling Perspective. *Current Directions in Psychological Science*, *27*(1), 56–63. https://doi.org/10.1177/0963721417727520

Goswami, U. (2019). Speech rhythm and language acquisition: an amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences*, 1–12. https://doi.org/10.1111/nyas.14137

Goswami, U., Power, A. J., Lallier, M., & Facoetti, A. (2014). Oscillatory ' Temporal Sampling ' and Developmental Dyslexia : Towards an Over-Arching Theoretical Framework. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2014.00904

Gourévitch, B., Martin, C., Postal, O., & Eggermont, J. J. (2020). Oscillations in the auditory system and their possible role. *Neuroscience and Biobehavioral Reviews*, *113*(December 2019), 507–528. https://doi.org/10.1016/j.neubiorev.2020.03.030

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review.

Neuroscience and Biobehavioral Reviews, *86*(December 2017), 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Hämäläinen, J. a, Rupp, A., Soltész, F., Szücs, D., & Goswami, U. (2012). Reduced phase locking to slow amplitude modulation in adults with dyslexia: an MEG study. *NeuroImage*, *59*(3), 2952–2961. https://doi.org/10.1016/j.neuroimage.2011.09.075

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. https://doi.org/10.1037/0033-295X.111.3.662

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2–3), 61–83; discussion 83-135. https://doi.org/10.1017/S0140525X0999152X

Hjetland, H. N., Lervåg, A., Lyster, S. A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology*, *111*(5), 751–763. https://doi.org/10.1037/edu0000321

Hoeft, F., McCandliss, B. D., Black, J. M., Gantman, A., Zakerani, N., Hulme, C., Lyytinen, H., Whitfield-Gabrieli, S., Glover, G. H., Reiss, A. L., & Gabrieli, J. D. E. (2011). Neural systems predicting long-term outcome in dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(1), 361–366. https://doi.org/10.1073/pnas.1008950108

Holliman, A. J., Gutiérrez Palma, N., Critten, S., Wood, C., Cunnane, H., & Pillinger, C. (2017). Examining the independent contribution of prosodic sensitivity to word reading and spelling in early readers. *Reading and Writing*, *30*(3), 509–521. https://doi.org/10.1007/s11145-016-9687-z

Holliman, A. J., Williams, G. J., Mundy, I. R., Wood, C., Hart, L., & Waldron, S. (2014). Beginning to disentangle the prosody-literacy relationship: A multi-component measure of prosodic sensitivity. *Reading and Writing*, *27*(2), 255–266. https://doi.org/10.1007/s11145-013-9443-6

Holliman, A., Mundy, I. R., Wade-Woolley, L., Wood, C., & Bird, C. (2017). Prosodic awareness and childrens' multisyllabic word reading. *Educational Psychology*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Holopainen, L., Ahonen, T., & Lyytinen, H. (2014). Predicting delay in reading achievement in a highly transparent language. *Journal of Learning Disabilities*, *34*(5), 401–413. https://doi.org/10.1177/002221940103400502

Huettig, F., Lachmann, T., Reis, A., & Petersson, K. M. (2018). Distinguishing cause from effect–many deficits associated with developmental dyslexia may be a consequence of reduced and suboptimal reading experience. *Language, Cognition and Neuroscience*, *33*(3), 333–350. https://doi.org/10.1080/23273798.2017.1348528

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The Causal Role of Phoneme Awareness and Letter-Sound Knowledge in Learning to Read: Combining Intervention Studies With Mediation Analyses. *Psychological Science*, *23*(6), 572–577. https://doi.org/10.1177/0956797611435921

Hulme, C., & Snowling, M. J. (2013). Learning to Read: What We Know and What We Need to Understand Better. *Child Development Perspectives*, *7*(1), 1–5. https://doi.org/10.1111/cdep.12005

Hulme, Charles, Snowling, M., Caravolas, M., & Carroll, J. (2005). Phonological Skills Are (Probably) One Cause of Success in Learning to Read: A Comment on Castles and Coltheart. *Scientific Studies of Reading*, *9*(4), 351–365.

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., & Giraud, A. L. (2015). Speech encoding by coupled cortical theta and gamma oscillations. *ELife*, *4*(MAY), 1–45. https://doi.org/10.7554/eLife.06213

INEEd. (2016). *Revisión de políticas para mejorar la efectividad del uso de los recursos educativos*. Informe país –

Uruguay, mimeo.

Ingram, J. C. L. (2007). *Neurolinguistics*. Cambridge University Press.

Jarmulowicz, L., Hay, S. E., Taran, V. L., & Ethington, C. A. (2008). Fitting derivational morphophonology into a developmental model of reading. *Reading and Writing*, *21*(3), 275–297. https://doi.org/10.1007/s11145-007-9073-y

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *NeuroImage*, *20*(2), 693–712. https://doi.org/10.1016/S1053-8119(03)00343-4

Junghöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, *37*(4), 523–532. https://doi.org/10.1017/S0048577200980624

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology*, *39*(3–4), 159–207. https://doi.org/10.1006/cogp.1999.0716

Kim, Y. S., Petscher, Y., Foorman, B. R., & Zhou, C. (2010). The Contributions of Phonological Awareness and Letter-Name Knowledge to Letter-Sound Acquisition-A Cross-Classified Multilevel Model Approach. *Journal of Educational Psychology*, *102*(2), 313–326. https://doi.org/10.1037/a0018449

Kirschner, S., & Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, *102*(3), 299–314. https://doi.org/10.1016/j.jecp.2008.07.005

Knoop-van Campen, C. A. N., Segers, E., & Verhoeven, L. (2018). How phonological awareness mediates the relation between working memory and word reading efficiency in children with dyslexia. *Dyslexia*, *May 2016*, 156–169. https://doi.org/10.1002/dys.1583

Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning Theory and Assessment of Reading Fluency: Automaticity, Prosody, and Definitions of Fluency. *Reading Research Quarterly*, *45*(2), 230–251. https://doi.org/10.1598/rrq.45.2.4

Kuhn, M. R., & Stahl, S. a. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, *95*(1), 3–21. https://doi.org/10.1037/0022-0663.95.1.3

Lallier, M., Molinaro, N., Lizarazu, M., Bourguignon, M., & Carreiras, M. (2016). Amodal atypical neural oscillatory activity in dyslexia: A cross-linguistic perspective. *October*, 1–57. https://doi.org/10.1095/biolreprod.109.079590

Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G., Parrila, R., & Georgiou, G. K. (2019). Phonological Awareness and Rapid Automatized Naming as Longitudinal Predictors of Reading in Five Alphabetic Orthographies with Varying Degrees of Consistency. *Scientific Studies of Reading*, *23*(3), 220–234. https://doi.org/10.1080/10888438.2018.1510936

Lehongre, K., Morillon, B., Giraud, A.-L., & Ramus, F. (2013). Impaired auditory sampling in dyslexia: further evidence from combined fMRI and EEG. *Frontiers in Human Neuroscience*, *7*(August), 454. https://doi.org/10.3389/fnhum.2013.00454

Lehongre, K., Ramus, F., Villiermet, N., Schwartz, D., & Giraud, A.-L. (2011). Altered low-γ sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron*, *72*(6), 1080–1090. https://doi.org/10.1016/j.neuron.2011.11.002

Lenth, R. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://cran.r-project.org/package=emmeans

Leong, V., & Goswami, U. (2014). Assessment of rhythmic entrainment at multiple timescales in dyslexia: evidence for disruption to syllable timing. *Hearing Research*, *308*, 141–161. https://doi.org/10.1016/j.heares.2013.07.015

Leong, V., Hämäläinen, J., Soltész, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language*, *64*(1), 59–73. https://doi.org/10.1016/j.jml.2010.09.003

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, *18*(2), 201–212. https://doi.org/10.1016/0022-0965(74)90101-5

Lizarazu, M., Lallier, M., Molinaro, N., Bourguignon, M., Paz-Alonso, P. M., Lerma-Usabiaga, G., & Carreiras, M. (2015). Developmental evaluation of atypical auditory sampling in dyslexia: Functional and structural evidence. *Human Brain Mapping*, *36*(12), 4986–5002. https://doi.org/10.1002/hbm.22986

Lizarazu, M., Lalllier, M., Molinaro, N., Bourguignon, M., Paz-Alonso, P., Lerma-Usabiaga, G., & Carreiras, M. (2015). Developmental evaluation of atypical auditory sampling in dyslexia: Functional and structural evidence. *Human Brain Mapping*.

Lizarazu, M., Scotto di Covella, L., van Wassenhove, V., Rivière, D., Mizzi, R., Lehongre, K., Hertz-Pannier, L., & Ramus, F. (2021). Neural entrainment to speech and nonspeech in dyslexia: conceptual replication and extension of previous investigations. *Cortex*, *137*(February), 160–178. https://doi.org/10.1016/j.cortex.2020.12.024

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of Emergent Literacy and Early Reading Skills in Preschool Children : Evidence From a Latent-Variable Longitudinal Study. *Developmental Psychology*, *36*(5), 596–613. https://doi.org/10.1037//OOI2-1649.36.5.596

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Lundetræ, K., & Thomson, J. M. (2018). Rhythm production at school entry as a predictor of poor reading and spelling at the end of first grade. *Reading and Writing*, *31*(1), 215–237. https://doi.org/10.1007/s11145-017-9782-9

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

Lyytinen, H., Erskine, J., Hämäläinen, J., Torppa, M., & Ronimus, M. (2015). Dyslexia—Early Identification and Prevention: Highlights from the Jyväskylä Longitudinal Study of Dyslexia. *Current Developmental Disorders Reports*, *2*(4), 330–338. https://doi.org/10.1007/s40474-015-0067-1

Lyytinen, H., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A.-M., & Lyytinen, P. (2006a). Trajectories of Reading Development: A Follow-up From Birth to School Age of Children With and Without Risk for Dyslexia. *Merrill-Palmer Quarterly*, *52*(3), 514–546. https://doi.org/10.1353/mpq.2006.0031

Lyytinen, H., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A. M., & Lyytinen, P. (2006b). Trajectories of reading development: A follow-up from birth to school age of children with and without risk for dyslexia. *Merrill-Palmer Quarterly*, *52*(3), 514–546. https://doi.org/10.1353/mpq.2006.0031

Mann, V., & Wimmer, H. (2002). Phoneme awareness and pathways into literacy: A comparison of German and American children. *Reading and Writing*, *15*(7–8), 653–682. https://doi.org/10.1023/A:1020984704781

Martin, S. D., & Shapiro, E. (2011). Examining the accuracy of teachers' judgments of dibels performance. *Psychology in the Schools*, *74*(4), 274–283. https://doi.org/10.1002/pits

Martinez Perez, T., Majerus, S., & Poncelet, M. (2012). The contribution of short-term memory for serial order to early reading acquisition: Evidence from a longitudinal study. *Journal of Experimental Child Psychology*, *111*(4), 708–723. https://doi.org/10.1016/j.jecp.2011.11.007

Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The Random Forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading*, *20*(1), 20–33. https://doi.org/10.1080/10888438.2015.1107073

Maurer, U., Bucher, K., Brem, S., Benz, R., Kranz, F., Schulz, E., van der Mark, S., Steinhausen, H.-C., & Brandeis, D. (2009). Neurophysiology in preschool improves behavioral prediction of reading ability throughout primary school. *Biological Psychiatry*, *66*(4), 341–348. https://doi.org/10.1016/j.biopsych.2009.02.031

McBride-Chang, C. (1995). What Is Phonological Awareness? *Journal of Educational Psychology*, *87*(2), 179–192. https://doi.org/10.1037/0022-0663.87.2.179

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*(7), 293–299. https://doi.org/10.1016/S1364-6613(03)00134-7

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Seguí, J. (1981). The syllable's role in speech segmentation. *Journal of Verval Learning and Verbal Behavious*, *20*(3), 298–305.

Melby-Lervåg, M., Lyster, S.-A., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*(March 2019), 104092. https://doi.org/10.1016/j.jml.2020.104092

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609–2621. https://doi.org/10.1111/ejn.13748

Meyer, L., Sun, Y., & Martin, A. E. (2019). Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, *0*(0), 1–11. https://doi.org/10.1080/23273798.2019.1693050

Ministerio de Salud. (2017). *Detección precoz de hipoacusia congenita*.

Molfese, D. L. (2000). Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain and Language*, *72*(3), 238–245. https://doi.org/10.1006/brln.2000.2287

Molinaro, N., & Lizarazu, M. (2017). Delta(but not theta)-band cortical entrainment involves speech-specific processing. *European Journal of Neuroscience*, *53*(9), 21–25. http://www.elsevier.com/locate/scp

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping*, *2783*(April), 2767–2783. https://doi.org/10.1002/hbm.23206

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Streiftau, S., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Démonet, J.-F., Longeras, E., Valdois, S., George, F., … Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, *29*, 65–77. https://doi.org/10.1016/j.learninstruc.2013.09.003

Morillon, B., Lehongre, K., Frackowiak, R. S. J., Ducorps, A., Kleinschmidt, A., Poeppel, D., & Giraud, A. L.

(2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(43), 18688–18693. https://doi.org/10.1073/pnas.1007189107

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology*, *40*(5), 665–681. https://doi.org/10.1037/0012-1649.40.5.665

Norton, E. S., & Wolf, M. (2012). Rapid Automatized Naming (RAN) and Reading Fluency: Implications for Understanding and Treatment of Reading Disabilities. *Annual Review of Psychology*, *63*(1), 427–452. https://doi.org/10.1146/annurev-psych-120710-100431

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*. http://fieldtrip.fcdonders.nl/_media/oostenveld_et_al_2011_computational_intelligence_and_neuroscience.pdf

Ortiz Barajas, M. C., Guevara, R., & Gervain, J. (2021). The origins and development of speech envelope tracking during the first months of life. *Developmental Cognitive Neuroscience*, *48*. https://doi.org/10.1016/j.dcn.2021.100915

Ozernov-Palchik, O., & Gaab, N. (2016). Tackling the “dyslexia paradox”: Reading brain and behavior for early markers of developmental dyslexia. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(2), 156–176. https://doi.org/10.1002/wcs.1383

Ozernov-Palchik, O., Norton, E. S., Sideridis, G., Beach, S. D., Wolf, M., Gabrieli, J. D. E., & Gaab, N. (2017). Longitudinal stability of pre-reading skill profiles of kindergarten children: implications for early screening and theories of reading. *Developmental Science*, *20*(5), 1–18. https://doi.org/10.1111/desc.12471

Ozernov-Palchik, O., & Patel, A. D. (2018). Musical rhythm and reading development: Does beat processing matter? *Annals of the New York Academy of Sciences*, 1–10. https://doi.org/10.1111/nyas.13853

Ozernov-Palchik, O., Wolf, M., & Patel, A. D. (2018). Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *Journal of Experimental Child Psychology*, *167*, 354–368. https://doi.org/10.1016/j.jecp.2017.11.009

Papadopoulos, T. C., Spanoudis, G., & Kendeou, P. (2009). The Dimensionality of Phonological Abilities in Greek. *Reading Research Quarterly*, *44*(2), 127–143. https://doi.org/10.1598/rrq.44.2.2

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Current Biology*, *25*(12), 1649–1653. https://doi.org/10.1016/j.cub.2015.04.049

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*(SEP). https://doi.org/10.3389/fpsyg.2012.00320

Peirce, J. W. (2008). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10. https://doi.org/10.3389/neuro.11.010.2008

Peng, P., Fuchs, D., Fuchs, L. S., Elleman, A. M., Kearns, D. M., Gilbert, J. K., Compton, D. L., Cho, E., & Patton, S. (2019). A Longitudinal Analysis of the Trajectories and Predictors of Word Reading and Reading Comprehension Development Among At-Risk Readers. *Journal of Learning Disabilities*, *52*(3), 195–208. https://doi.org/10.1177/0022219418809080

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., Leopold, D. R., Samuelsson, S., Byrne, B., Willcutt, E. G., & Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology*, *121*(1), 212–224. https://doi.org/10.1037/a0025823

Petscher, Y., & Koon, S. (2020). Moving the Needle on Evaluating Multivariate Screening Accuracy. *Assessment for Effective Intervention*, *45*(2), 83–94. https://doi.org/10.1177/1534508418791740

Piasta, S. B., & Wagner, R. K. (2010). Learning letter names and sounds: Effects of instruction, letter type, and phonological processing skill. *Journal of Experimental Child Psychology*, *105*(4), 324–344. https://doi.org/10.1016/j.jecp.2009.12.008

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, *21*(6), 322–334. https://doi.org/10.1038/s41583-020-0304-4

Power, A. J., Colling, L. J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, *160*, 1–10. https://doi.org/10.1016/j.bandl.2016.06.006

Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in Human Neuroscience*, *7*, 777. https://doi.org/10.3389/fnhum.2013.00777

Prinsloo, K. D., & Lalor, E. C. (2020). General auditory and speech-specific contributions to cortical envelope tracking revealed using auditory chimeras. *BioRxiv*. https://doi.org/10.1101/2020.10.21.348557

Protopapas, A. (2019). Evolving Concepts of Dyslexia and Their Implications for Research and Remediation. *Frontiers in Psychology*, *10*, 1–10. https://doi.org/10.3389/fpsyg.2019.02873

Protopapas, A., Gerakaki, S., & Alexandri, S. (2006). Lexical and default stress assignment in reading Greek. *Journal of Research in Reading*, *29*(4), 418–432. https://doi.org/10.1111/j.1467-9817.2006.00316.x

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A., Tolvanen, A., Torppa, M., & Lyytinen, H. (2008). Developmental links of very early phonological and language skills to second grade reading outcomes: strong to accuracy but only minor to fluency. *Journal of Learning Disabilities*, *41*(4), 353–370. https://doi.org/10.1177/0022219407311747

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Poikkeus, A., Tolvanen, A., & Torppa, M. (2007). Very early phonological and language skills : estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry*, *9*, 923–931. https://doi.org/10.1111/j.1469-7610.2007.01763.x

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Ramus, F. (2004). Neurobiology of dyslexia: A reinterpretation of the data. *Trends in Neurosciences*, *27*(12), 720–726. https://doi.org/10.1016/j.tins.2004.10.004

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology (2006)*, *61*(1), 129–141. https://doi.org/10.1080/17470210701508822

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive Sensing of Periodic and Aperiodic Auditory Patterns. *Trends in Cognitive Sciences*, *22*(10), 870–882. https://doi.org/10.1016/j.tics.2018.08.003

Ríos-López, P., Molinaro, N., Bourguignon, M., & Lallier, M. (2020). Development of neural oscillatory activity in response to speech in children from 4 to 6 years old. *Developmental Science*, *23*(6), 1–16. https://doi.org/10.1111/desc.12947

Rios Lopez, P. (2018). *Investigation of the Development of Neural and Behavioural Auditory Rhythmic Sensitivity and of its Contribution to Reading Acquisition* (Issue September).

Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling. *Journal Of Statistical Software*, *17*(5), 1–25. http://rsirt.googlecode.com/files/ltmPackage.pdf

Runge, T. J., & Watkins, M. W. (2006). The structure of phonological awareness among kindergarten students. *School Psychology Review*, *35*(3), 370–386.

Scarborough, H. S., & Brady, S. A. (2002). Toward a Common Terminology for Talking about Speech and Reading: A Glossary of the "Phon" Words and Some Related Terms. *Journal of Literacy Research*, *34*(3), 299–336. https://doi.org/10.1207/s15548430jlr3403_3

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten Prediction of Reading Skills: A Longitudinal Comparative Analysis. *Journal of Educational Psychology*, *96*(2), 265–282. https://doi.org/10.1037/0022-0663.96.2.265

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology*, *91*(3), 439–449. https://doi.org/10.1037//0022-0663.91.3.439

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin and Review*, *22*(6), 1614–1629. https://doi.org/10.3758/s13423-015-0835-2

Schmitterer, A. M. A., & Schroeder, S. (2019). Effects of reading and spelling predictors before and after school entry: Evidence from a German longitudinal study. *Learning and Instruction*, *59*(2019), 46–53. https://doi.org/10.1016/j.learninstruc.2018.09.005

Schroeder, C., & Lakatos, P. (2010). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 1–16. https://doi.org/10.1016/j.tins.2008.09.012.Low-frequency

Seidenberg, M. S. (2017). *Language at the speed of sight: how we read, why so many can't, and what can be done about it*. New York, NY: Basic Books.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology (London, England: 1953)*, *94*(Pt 2), 143–174. https://doi.org/10.1348/000712603321661859

Share, D. L. (2008). *On the Anglocentricities of Current Reading Research and Practice: The Perils of Overreliance on an " Outlier " Orthography*. *134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Smalle, E. H. M., Szmalec, A., Bogaerts, L., Page, M. P. A., Narang, V., Misra, D., Araújo, S., Lohagun, N., Khan, O., Singh, A., Mishra, R. K., & Huettig, F. (2019). Literacy improves short-term serial recall of spoken verbal but not visuospatial items – Evidence from illiterate and literate adults. *Cognition*, *185*(January 2018), 144–150. https://doi.org/10.1016/j.cognition.2019.01.012

Solheim, O. J., Torppa, M., & Uppstad, P. H. (2020). Screening for Slow Reading Acquisition in Norway and Finland – a Quest for Context Specific Predictors Screening for Slow Reading Acquisition in Norway and. *Scandinavian Journal of Educational Research*, *0*(0), 1–17. https://doi.org/10.1080/00313831.2020.1739130

Sprenger-Charolles, L., Siegel, L. S., Jiménez, J. E., & Ziegler, J. C. (2011). Prevalence and Reliability of Phonological, Surface, and Mixed Profiles in Dyslexia: A Review of Studies Conducted in Languages Varying in Orthographic Depth. *Scientific Studies of Reading*, *15*(6), 498–521. https://doi.org/10.1080/10888438.2010.524463

Stanovich, K. E. (1986). Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. *Reading Research Quarterly*, *21*(4), 360–407. https://doi.org/10.1598/RRQ.21.4.1

Steinbrink, C., Knigge, J., Mannhaupt, G., Sallat, S., & Werkle, A. (2019). Are temporal and tonal musical skills related to phonological awareness and literacy skills? - Evidence from two cross-sectional studies with children from different age groups. *Frontiers in Psychology*, *10*(MAR), 1–16. https://doi.org/10.3389/fpsyg.2019.00805

Sunde, K., Furnes, B., & Lundetræ, K. (2019). Does Introducing the Letters Faster Boost the Development of Children's Letter Knowledge, Word Reading and Spelling in the First Year of School? *Scientific Studies of Reading*, *0*(0), 1–18. https://doi.org/10.1080/10888438.2019.1615491

The Mathworks Inc. (2018). MATLAB R2018a. In *www.mathworks.com/products/matlab*. https://doi.org/2016-11-26

Thompson, P. a., Hulme, C., Nash, H. M., Gooch, D., Hayiou-Thomas, E., & Snowling, M. J. (2015). Developmental dyslexia: predicting individual risk. *Journal of Child Psychology and Psychiatry*, *56*(9), 976–987. https://doi.org/10.1111/jcpp.12412

Tierney, A., & Kraus, N. (2013). The Ability to Move to a Beat Is Linked to the Consistency of Neural Responses to Sound. *Journal of Neuroscience*, *33*(38), 14981–14988. https://doi.org/10.1523/JNEUROSCI.0612-13.2013

Tierney, A. T., & Kraus, N. (2013). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain and Language*, *124*(3), 225–231. https://doi.org/10.1016/j.bandl.2012.12.014

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, *27*(5). https://doi.org/10.1177/002221949402700503

Treiman, R., & Kessler, B. (2004). The Role of Letter Names in the Acquisition of Literacy. *Advances in Child Development and Behavior*, *31*(C), 105–135. https://doi.org/10.1016/S0065-2407(03)31003-1

Treiman, R., Tincoff, R., & Richmond-Welty, E. D. (1996). Letter names help children to connect print and speech. *Developmental Psychology*, *32*(3), 505–514. https://doi.org/10.1037/0012-1649.32.3.505

Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. a, & Eden, G. F. (2003). Development of neural mechanisms for reading. *Nature Neuroscience*, *6*(7), 767–773. https://doi.org/10.1038/nn1065

Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. (2010). Cognitive Development of Fluent Word Reading Does Not Qualitatively Differ Between Transparent and Opaque Orthographies. *Journal of Educational Psychology*, *102*(4), 827–842. https://doi.org/10.1037/a0019465

Van Bergen, E., De Jong, P. F., Regtvoort, A., Oort, F., Van Otterloo, S., & Van Der Leij, A. (2011). Dutch children at family risk of dyslexia: Precursors, reading development, and parental effects. *Dyslexia*, *17*(1), 2–18. https://doi.org/10.1002/dys.423

Van Orden, G. C., & Kloos, H. (2005). Phonology and reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 1–46).

Vanvooren, S., Poelmans, H., Hofmann, M., Ghesquière, P., & Wouters, J. (2014). Hemispheric asymmetry in auditory processing of speech envelope modulations in prereading children. *Journal of Neuroscience*, *34*(4), 1523–1529. https://doi.org/10.1523/JNEUROSCI.3209-13.2014

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades ? *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40.

Vellutino, F. R., & Scanlon, D. M. (2013). Phonological Coding , Phonological Awareness , and Reading Ability : Evidence from a Longitudinal and Experimental Study. *Merrill-Palmer Quarterly*, *33*(3), 321–363.

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing*, *21*(4), 437–480. https://doi.org/10.1007/s11145-007-9098-2

Verhoeven, L., & Keuning, J. (2017). The Nature of Developmental Dyslexia in a Transparent Orthography.

Scientific Studies of Reading, 1–17. https://doi.org/10.1080/10888438.2017.1317780

Vloedgraven, J. M. T., & Verhoeven, L. (2007). Screening of phonological awareness in the early elementary grades: An IRT approach. *Annals of Dyslexia*, *57*(1), 33–50. https://doi.org/10.1007/s11881-007-0001-2

Vloedgraven, J. M. T., & Verhoeven, L. (2009). The nature of phonological awareness throughout the elementary grades: An item response theory perspective. *Learning and Individual Differences*, *19*(2), 161–169. https://doi.org/10.1016/j.lindif.2008.09.005

Vousden, J. I., Ellefson, M. R., Solity, J., & Chater, N. (2011). Simplifying reading: Applying the simplicity principle to reading. *Cognitive Science*, *35*(1), 34–78. https://doi.org/10.1111/j.1551-6709.2010.01134.x

Wanzek, J., & Vaughn, S. (2007). Research’Based Implications From Extensive Early Reading Interventions. *School Psychology Review*, *36*(4), 541–561.

Wechsler, D. (2001). *WPPSI-III: Escala de Inteligencia de Wechsler para Preescolar y Primaria*. TEA.

White, S., Milne, E., Rosen, S., Hansen, P., Swettenham, J., Frith, U., & Ramus, F. (2006). The role of sensorimotor impairments in dyslexia: A multiple case study of dyslexic children. *Developmental Science*, *9*(3), 237–255. https://doi.org/10.1111/j.1467-7687.2006.00483.x

Wood, C., & Connelly, V. (2009). Phonological awareness beyond phonemes. In *Contemporary Perspectives on Reading and Spelling*. https://doi.org/10.4324/9780203877838

Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences*, *111*(40), 14559–14564. https://doi.org/10.1073/pnas.1406219111

Yopp, H. K. (1988). The Validity and Reliability of Phonemic Awareness Tests. *Reading Research Quarterly*, *23*(2), 159. https://doi.org/10.2307/747800

Ziegler, J. C., Bertrand, D., Toth, D., Csepe, V., Reis, a., Faisca, L., Saine, N., Lyytinen, H., Vaessen, a., & Blomert, L. (2010). Orthographic Depth and Its Impact on Universal Predictors of Reading: A Cross-Language Investigation. *Psychological Science*, *21*(4), 551–559. https://doi.org/10.1177/0956797610363406

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

Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language*, *93*(3), 369–373. https://doi.org/10.1016/j.bandl.2004.10.010

Appendix A Phonological awareness psychometrics

In the present section we examine the psychometric properties of the phonological assessment battery in Lexiland. First, we describe the available standardized assessment batteries of phonological awareness skills in Spanish, and the rationale for developing our own. Then, we present with the psychometric properties of our phonological awareness battery, including its reliability, dimensionality, and validity.

A.1 Introduction

A.1.1 Available standardized instruments for phonological awareness assessment in Spanish

In United States, in a seminal study, Anthony and colleagues (2011) used two tests from the Spanish Preschool Comprehensive Test of Phonological and Print Processing (Lonigan & Farver, 2002) to assess PA's dimensionality and sequence of development. While the study has had a moderate impact (60 citations to date), the battery is unpublished and has not been widely used; all the citations available come from this same research group, at Florida State University at the time. Additional tools for Spanish-speaking populations of children in the US from related and unrelated research groups follow the same pattern, none of them provide norming data or detailed descriptions of item design (C-PALLS: Landry, Anthony, Swank, & Monseque-Bailey, 2009; CFE: Riccio et al., 2001, GRTR-S: http://www.getreadytoread.org). Probably as a consequence of the lack of data, most of them are used by the same research group that designed it. Fortunately, a recent article addressed this gap by designing, norming and publishing a novel phonological awareness test for this population (Wackerle-Hollman et al., 2019). In Spain, a prolific line of research by Sylvia Defior and colleagues routinely use ad-hoc tests of phonological awareness (see for example Calet, Gutiérrez-Palma, Simpson, González-Trujillo, & Defior, 2015; Defior, Serrano, & Marín-Cano, 2008). The same is the case for other research groups in Spain (Carrillo, 1994; Cuetos, Martínez-García, & Suárez-Coalla, 2018). Although we know of one normed test of phonological awareness for the Spanish population (Jiménez & Ortiz, 1995), it is not readily available. In Argentina, Chile and Uruguay, ad-hoc tests are also the norm. In Argentina, the most comprehensive assessment that we know of consists of an unpublished doctoral thesis, but it is aimed at school-aged children and not pre-schoolers (Pearson, 2012). A more modest attempt is found in Manrique (Manrique & Graminga, 1984; Manrique & Signorini, 1994). So is the case in Chile (Guardia Gutiérrez, 2003, 2010). In Uruguay, the only published work relative to phonological awareness assessment in pre-schoolers that we know of, employs *prueba de segmentación lingüística*, PSL (Cuadro & Trías, 2008). Finally, cross-linguistic studies of phonological awareness and reading involving Spanish-speaking populations also employ ad-hoc assessments (Caravolas, Lervåg, Defior, Seidlová Málková, & Hulme, 2013; Caravolas et al., 2012; Duncan et al., 2013; Moll et al., 2014).

In summary, while phonological awareness is regarded as a central ability in reading acquisition, there are currently almost no readily available normed tests for Spanish-speaking populations. In the present work, we developed our own assessment of phonological awareness abilities. In the present chapter we describe its main psychometric properties. Our aim is to validate the quality of PA assessment relevant for the research questions present in the thesis. Therefore, we focus on the properties of K5 assessment. A more complete psychometric analysis, though urgent, is beyond the scope of the present thesis. To fulfil this aim we study the batterie's reliability, dimensionality (through factor analysis), and validity (through a criterion measure). In the following section we examine the battery's reliability, dimensionality, and validity.

A.1.2 Sample and design

Sampling comprised 26 public schools in Montevideo (Uruguay). All schools were above the fourth quintile in socioeconomic status (Q4 = 9 schools, Q5 = 17 schools), according to the public school system rating (Administración Nacional de Educación Pública, ANEP). Schools were either part-time or full-time. All children attending K5 level at Time 1 (821 children) were invited to take part in the study. Only those whose parents signed the consent form finally took part. Sample size at Time 1 included 616 (75%) children. At Time 2, 397 (64.4 %) out of the original 616 children continued in the study. According to the data available in the public-school system database (GURI), 76% of the children continued in G1 at the same school where they had attended K5, 5% moved to private schools and 13 % switched between public schools. At Time 2 one of the schools dropped out of the study for scheduling reasons (2.5% of children). The remaining 2.5 % could not be tracked (most of them due to a mismatch between their ID number in our database and the one in GURI). At Time 3, all children that had taken part at Time 1 or Time 2 and that were still attending any one of the 26 participating schools were invited to continue the study, except for 5 schools that could not continue for scheduling reasons (92 children). At Time 3, 250 children continued in the study (62.9 % of Time 2 sample, 40.5 % of Time 1 sample). We do not have access to the mobility occurring between Time 1 and Time 2 thus we cannot describe the reasons for the dropout.

Time 1 data collection took place in the second trimester of the school year, between June and August 2016; Time 2 and Time 3 data collection took place in the last trimester of the school year, between October and December 2017 and 2018 (in Uruguay the academic year starts in March and ends in December).

Children were assessed at their School, in groups of 4 to 5. Each child was assessed in 4-5 sessions, approximately 20 minutes each in Time 1 and Time 2, and 1 session of 20 minutes at Time 3 (only reading measures were included at this timepoint). Two research assistants monitored task performance and were available to clarify instructions on demand.

A.1.3 PA tasks

All tasks were presented using a tablet-based App -*Lexiland*- developed by the research team (Figure 2.1)In order to increase children’s motivation and engagement in autonomous play, tasks where embedded in a videogame-like ludic narrative, with a main character and rewards for task completion. All tasks consisted of 2 to 3 example trials, 4 to 5 practice trials with feedback, followed by test trials without feedback. Effort was made to avoid the need to obtain verbal responses, in order to automatize data collection and processing. Thus, verbal responses were replaced by multiple choice items when possible (except for the Reading and RAN tasks). Instructions and auditory stimuli were pre-recorded and presented via headphones. Response times and errors were recorded in all tasks.

Figure A.1 Lexiland videogame screenshots. Left to right, top to bottom: Segmentation, Blending, Onset matching and Rhyme, Letter Knowledge, RAN, Vocabulary, Verbal Short-term memory, non-verbal Short-term memory, IQ.

Phonological awareness was assessed through four tasks: segmentation, blending, onset matching and rhyme. For each task, two separate subtasks at the syllable and phoneme levels were presented (except for rhyming).

- Segmentation (22 syllabic items, 28 phonemic items): a word was presented aurally together with a picture of it, children were asked to segment a word in either syllables or phonemes. In order to avoid verbal responses, together with the picture of the word, illustrations of dices corresponding to number two to four for syllables, and three to five for phonemes appeared in the screen. The answer was given by tapping on the dice corresponding to the number of syllables or phonemes in the word. Within each grain size, items ranged from two to four syllables, and three to five phonemes. Within each length, approximately half of the items began with CV syllables, and half with CCV syllables.

- Blending (18 syllabic items, 16 phonemic items): children were asked to blend aurally presented syllabic or phonemic segments into a word. The answer was given by selecting one out of four pictures presenting the target word and three distractors (one semantically related, one phonologically related and one unrelated). Within each grain size, items ranged from two to four syllables, and four to six phonemes. Within each length, approximately half of the items began with CV syllables, and half with CCV syllables.
- Onset matching (27 syllabic items, 32 phonemic items) and rhyme (10 word items, 10 pseudowords items): children heard pairs of words (rhyme also included pseudowords) and saw pictures for each of them (except for pseudowords). They had to answer whether both words started with the same syllable or phoneme (isolation) or rhymed (rhyme). The answer was given by tapping on a tick or a cross on the screen. For onset matching, within each grain size, items ranged from two to three syllables, and four to six phonemes. Within each length, approximately half of the items began with CV syllables, and half with CCV syllables. For rhyme, all items had three syllables and a CV syllable structure.

A.2 Reliability: Item-total correlation and Cronbach's alpha reliability

Item total-correlations and Cronbach's alpha reliability were computed for each item and each task (Figure A.2 and Table A. 2. Fit indices for the four, five, eight and twelve factor solutions). As can be seen from the figure, segmentation phonemes shows the lowest item-total correlation, and blending syllables the highest. Overall, Cronbach's alpha's are moderate to high, except for the rhyme task. Finally, it is worth noting that average inter-item reliability is low, reflecting the fact that within each tasks, there were experimental manipulations such as increasing word length and syllabic complexity (from CV to CCV onset).

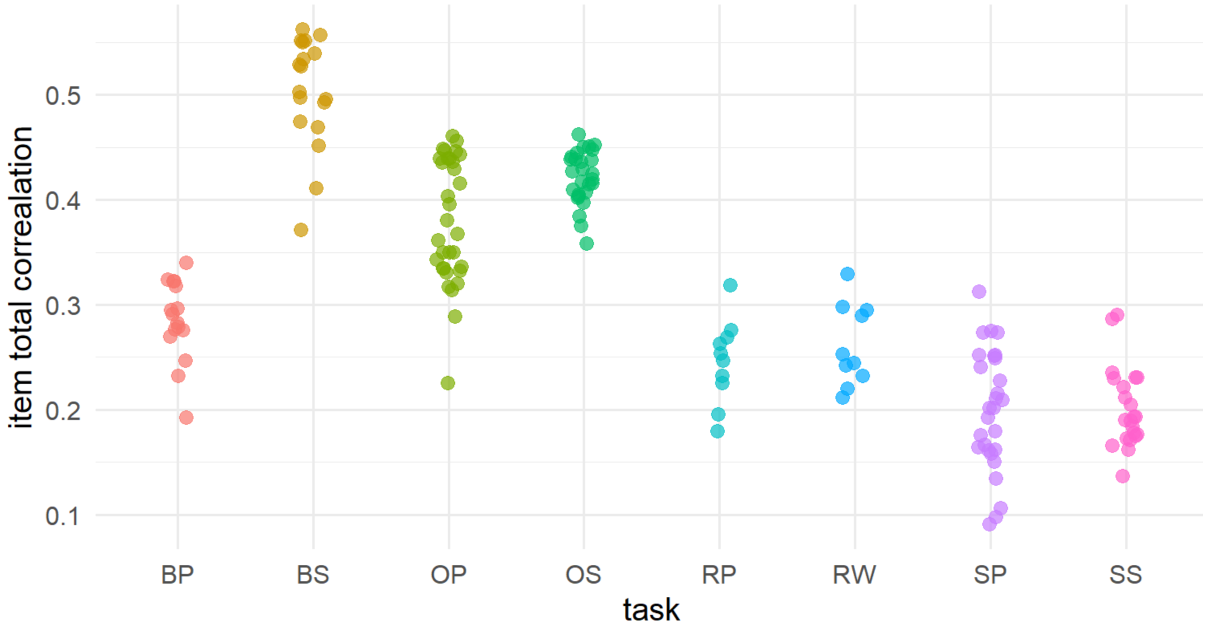


Figure A.2. Item-total correlation by task

Table A. 1. Reliability for each task

	BP	BS	SP	SS	OP	OS	RW	RPW
raw alpha	0.628	0.745	0.583	0.715	0.569	0.707	0.281	0.254
std alpha	0.630	0.759	0.585	0.714	0.566	0.699	0.281	0.254
lambda G6	0.629	0.762	0.647	0.749	0.673	0.753	0.333	0.317
avg r	0.096	0.149	0.048	0.102	0.039	0.079	0.038	0.033

Note: BP: blending phonemes; BS: blending syllables; SP: segmentation phonemes; SS: segmentation syllables; OP: onset-matching phonemes; OS: onset-matching syllables; RW: rhyme words; RPW: rhyme pseudo-words. Raw alpha: Cronbach's alpha based upon the covariances; Std alpha: Cronbach's alpha based upon the correlations; lambda G6: Guttman's Lambda 6 reliability; avg r: average inter-item correlation.

A.3 Dimensionality: Factor analysis

A.3.1 Procedure

Principal component (PC) and principal axis factors (FA) were extracted from children's raw K5 responses. Ordinary lest squares regression was used to find the minimum residual solution through the *fa.parallel* function of the psych package in R (Revelle, 2021). Tetrachoric correlations were used to account for the dichotomous nature of responses. For FA, oblimin and varimax rotations were considered in addition to the base solution without rotation.

In order to select number of factor/components, several indices were considered: i) parallel analysis (PC and FA), ii) Kaiser criterion (eigen values larger than 1), iii) scree plot, and iv) Bayes information criterion. Parallel analysis follows Horn's procedure (Horn, 1965). In brief, it compares PC and FA solutions to the solutions produced for two alternative datasets: one generated by a random matrix of univariate normal of data, and one generated by randomly resampling the real data. Eigen values are compared for the three datasets. Factors/Components with eigenvalues larger than those in the simulated/sampled dataset are retained, i.e., where scree plots cross-over.

For each solution, the following criteria were used for item exclusion: i) correlations lower than 0.25, ii) communalities smaller than 0.2 and loadings smaller than 0.4, iii) loadings larger than 0.3 in more than one factor. Items that fulfilled any one of the criteria were regarded as atypical. For each factor solution, model was re-fit iteratively until no item matched the exclusion criteria.

A.3.2 Results

The large number of items in the battery yielded very large number of factors according to all the criteria used: parallel PC suggested 16, parallel FA 22, Kaiser criterion 41 and BIC 17. Therefore, number of factors was estimated on the basis of strong theoretical accounts and by design. All following solutions pertain to the oblimin rotation, which is the one that yielded more interpretable results.

Initially, a 4-factor solution was tested, assuming each task should load into one separate factor. Loadings larger than 0.3 in the pattern matrix distribute as follows:

1. BS & BP
2. OP & OS matching onset
3. OP & OS; non-matching onset
4. SS

Sixty-one (37%) items were excluded in this solution. These include all items from the rhyme tasks (RW & RPW) and all items from the SP task. Additionally, 6 items from BP and 2 items from SS were also excluded. However, rhyme has been described as a separate phonological awareness ability in previous studies (Muter, Hulme, Snowling, & Taylor, 1998; Runge & Watkins, 2006), therefore, a 5-factor solution was estimated in trying to account for rhyme items.

In the 5-factor solution, loadings larger than 0.3 in the pattern matrix distribute as follows:

1. BP & BS
2. OP & OS matching onset
3. OP & OS; non-matching onset
4. RW & RPW
5. SS

In this solution, 35 items were excluded, including the whole segmentation phonemes task (28 items), six from blending phonemes, and one from segmentation syllables.

The fact that onset matching items load into two separate factors is somewhat surprising, especially considering that the fact that items load into each factor depending on item properties: items that begin with matching onsets load into one factor, and items with non-matching onset load into another factor. It is possible that FA is discovering the underlying structure of the onset matching items. However, it is also possible to explain this phenomenon in terms of careless responding. Careless responding has been mainly described in the personality research literature to account for the type of responses where the subject responds without considering the information contained in the item. This effect is often interpreted as a result of tiredness or lack of interest (Kam & Meyer, 2015). A similar related phenomenon is acquiescence, a subject's tendency to agree with whatever information is contained in a given item. In our dataset, we believe it can be explained in terms of children's inability to perform the task due to its difficulty, and therefore responding persistently with the same button (either yes or no), irrespective of item information. This would be specially aggravated for phoneme level assessments in the K5 timepoint. Though not systematically addressed, this behaviour was informally reported by research assistants and observed by CZ.

An 8-factor solution was also tested, in line with the theoretical account of 4 tasks with 2 levels each (syllables and phonemes for onset matching , blending and segmentation; and words and pseudowords for rhyme). In the 8-factor solution, loadings larger than 0.3 in the pattern matrix distribute as follows:

1. BS & BP
2. OP matching onset
3. OS matching onset
4. OP non-matching onset & OS non-matching onset
5. SS 2-syllable words
6. SS 3-syllable words
7. SP 6-phoneme words
8. RW & RPW

Thirty-eight items were excluded consisting of: all SP items with 3 to 5 phonemes (20 items) and 2 with 6 phonemes (items 25 and 26), 6 SS items (all with 4 syllables), 7 BP items and 3 RPW items.

As with the 5-factor solution, onset matching items load into two separate factors depending on item properties (matching/non-matching onsets). Additionally, SS shows a similar pattern depending on items properties, in this case, number of syllables.

In further trying to characterize the pattern of results of the available solutions, and following the best solution according to parallel PC analysis, a 16-factor solution was estimated. Loadings larger than 0.3 in the pattern matrix distribute as follows:

1. BS & BP
2. OS matching onset
3. OS non-matching onset
4. OP matching onset
5. OP non-matching onset
6. RW that rhyme & RPW that rhyme
7. RW that don't rhyme & RPW that don't rhyme
8. SS 2-syllable words
9. SS 3-syllable words
10. SS 4-syllable words
11. SP 3-phoneme words
12. SP 4-phoneme words
13. SP 5-phoneme words
14. SP 6-phoneme words
15. Discarded (BS 1 item)
16. Discarded (BP items with loading smaller than .4)

Sixteen items (10%) are excluded: 14 BP, 1 BS and 1 SP. Given that the final solution after items were discarded included 2 factors with no loadings, a 14-factor solution was re-fit. Again, one factor contained loadings from BP items which were smaller than .4, thus a 13-factor solution was re-fit. The same pattern was observed: one factor contained BP items with loadings smaller than .4. Thus, a 12-factor solution was fit. Loadings in the 12-factor solution were as follow:

1. BS & BP
2. OS matching onset
3. OS non-matching onset
4. OP matching onset
5. OP non-matching onset
6. RW that rhyme & RPW that rhyme

7. RW that don’t rhyme & RPW that don’t rhyme
8. SS 2-syllable words
9. SS 3-syllable words
10. SS 4-syllable words
11. SP 4-phoneme words
12. SP 5-phoneme words

Nineteen items were discarded: 8 BP and 11 SP (3 and 6-phoneme items).

A.3.3 Best solution

Factor loadings and a cropped scree plot for the 12-factor solution are displayed in Figure A.3 and Figure A.4. Fit indices were obtained for the solutions presented above up to 12 factors. All fit indices converge in favouring the 12-factor solution (Table A. 2. Fit indices for the four, five, eight and twelve factor solutions). For this solution, Kaiser-Meyer-Olkin statistic was 0.73, total Cronbach’s alpha reliability is 0.945, average factor Cronbach’s alpha reliability is 0.83 (min: 0.70, max: 0.93)

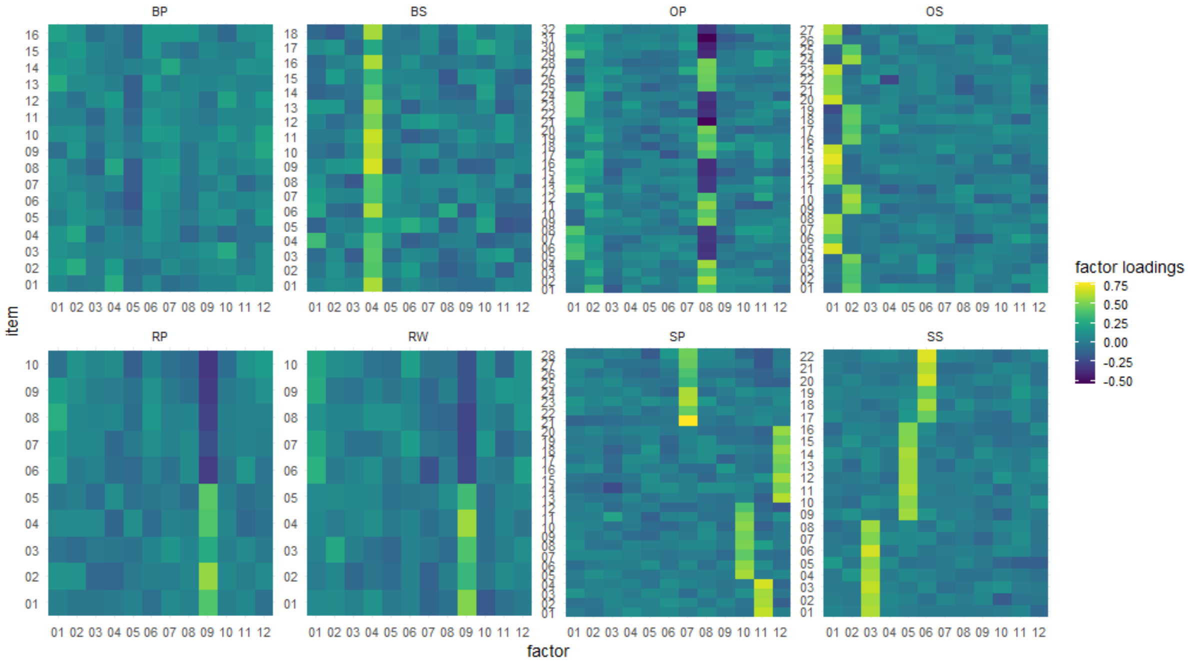


Figure A.3. Factors and factor loadings for each item in each task [notice factor number is arbitrary].

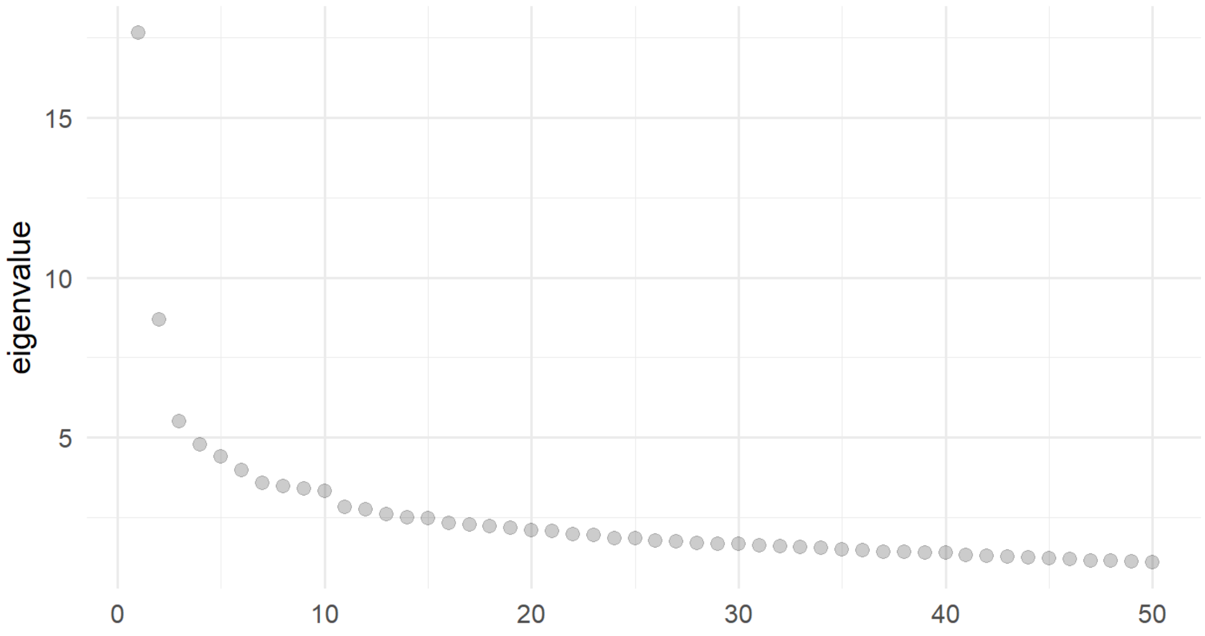


Figure A.4. Scree plot of first 50 eigenvalues ordered from maximum to minimum.

Table A. 2. Fit indices for the four, five, eight and twelve factor solutions

# factors	BIC	chi	dof	TLI	RMSEA
4	29382.32	75622.16	12557.00	0.86	0.11
5	10451.64	58589.37	12398.00	0.89	0.10
8	-11696.54	38968.22	11927.00	0.92	0.09
12	-25120.26	26658.36	11313.00	0.94	0.07

Note: BIC: Bayesian information criterion; chi: chi square statistic; dof: degrees of freedom; TLI: Tucker-Lewis index; RMSEA: root mean square error of approximation.

While model comparison favours a 12-factor solution, it is possible that all factors load into a second-order latent single factor of phonological awareness, as previous literature suggests (Anthony et al., 2002), which could be tested by confirmatory factor analysis via structural equation modelling. However, attempts to implement such a model in R, via *lavaan* or *sem* packages were fruitless. It was not possible to fit such a high dimensional model with any of the available functions. Some form of data reduction should therefore be employed. However, the aim of the present chapter was not to test the dimensionality of phonological awareness —however current the debate is— but to refine Lexiland PA as an assessment tool. Previous studies in English, Spanish, Greek and Dutch suggest phonological awareness is a unidimensional construct (Anthony et al., 2002, 2011; Papadopoulos, Spanoudis, & Kendeou, 2009; Vloedgraven & Verhoeven, 2007). Others, however, find its best described by a two-dimensional construct composed of intercorrelated factors representing phonemic units on one side, and syllabic and/or intra-syllabic units on the other (Muter et al., 1998). Even three-factor models have been described as adequate fits for the PA construct (Meira, Cadime, & Leopoldina Viana, 2019), though this latter referred to PA in Portuguese. Once again, differences in oral language properties, tasks used and, in many cases, small sample size, could probably explain the discrepancies. Crucially, the factor analysis faithfully captures the underlying constructs and manipulations present in the items by task and grain size, suggesting the instrument is accurately reflecting phonological awareness skills in prereading children.

A.4 Validity: criterion measure

In order to further examine the validity of the Lexiland phonological awareness battery, a new sample of 30 children were assessed with PCF10 (Cimino & Dalmás, 2001) by three trained speech-language pathologists. This assessment consists of 10 tasks with 4 items each. All tasks include an example trial and a practice trial with feedback. Children were assessed at their schools, individually. Additionally, children completed phonological awareness tasks from Lexiland, also individually at their schools, with the help of research assistants.

The battery consists of:

i. word length: the child hears two words of different length and has to indicate verbally which one is longer.
ii. syllable segmentation: the child hears a word and has to segment it into syllables.
iii. rhyme oddity: the child hears a target word and three words; he has to select the word that rhymes with the target.
iv. phoneme identification: the child hears a target word and three words; he has to select the word that begins with the same sound as the target.
v. phoneme blending: the child hears a word segmented in phonemes and has to orally blend them.
vi. phoneme segmentation: the child hears a word and has to segment it in syllables.
vii. final sound: the child hears a word and hast to produce a word that ends with the same sound.
viii. phoneme elision: the child hears a word, has to remove a target sound, and produce the resulting nonword.
ix. phoneme replacement: the child hears a word and has to replace a phoneme present in the word for a target phoneme.
x. phoneme addition: the child hears a word and has to add a phoneme to it.

Two children were removed due to missing data. Pearson correlation coefficient between mean accuracy scores for Lexiland and PCF10 was 0.72 ($p < 0.001$) for the complete assessment, and 0.74 ($p < 0.001$) for phoneme tasks only.

A.5 Conclusions

The aim of the present chapter was to assess the psychometric properties of Lexiland's phonological awareness assessment. Lexiland's PA assessment consists of 163 items distributed among eight tasks, including both syllable and phoneme level items. It was tested on approximately 600 K5 children of middle-income public schools in Montevideo. The results show both high reliability and validity, which posits Lexiland's phonological awareness battery as a robust assessment tool.

A.6 References

Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll, K., Phillips, B. M., & Cantor, B. G. (2002). Structure of preschool phonological sensitivity: Overlapping sensitivity to rhyme, words, syllables, and phonemes. *Journal of Experimental Child Psychology*, *82*(1), 65–92. https://doi.org/10.1006/jecp.2002.2677

Anthony, J. L., Williams, J. M., Durán, L. K., Gillam, S. L., Liang, L., Aghara, R., … Landry, S. H. (2011). Spanish Phonological Awareness: Dimensionality and Sequence of Development During the Preschool and Kindergarten Years. *Journal of Educational Psychology*, *103*(4), 857–876. https://doi.org/10.1037/a0025024

Calet, N., Gutiérrez-Palma, N., Simpson, I. C., González-Trujillo, M. C., & Defior, S. (2015). Suprasegmental Phonology Development and Reading Acquisition: A Longitudinal Study. *Scientific Studies of Reading*, *19*(1), 51–71. https://doi.org/10.1080/10888438.2014.976342

Caravolas, M., Lervåg, A., Defior, S., Seidlová Málková, G., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological Science*, *24*(8), 1398–1407. https://doi.org/10.1177/0956797612473122

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavsky, M., Onochie-Quintanilla, E., … Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Supplemental Material*, (March), 1–33.

Carrillo, M. (1994). Development of phonological awareness and reading acquisition - A study in spanish language. *Reading and Writing*, *6*(3), 279–298. https://doi.org/10.1007/BF01027086

Cimino, M., & Dalmás, L. (2001). Evaluación de la conciencia fonológica en niños de 4,2 a 8,11 años que concurren a una escuela privada de Montevideo. In *VII Congreso Latinoamericano de Neuropsicología*.

Cuadro, A., & Trías, D. (2008). Desarrollo de la conciencia fonémica : Evaluación de un programa de intervención. *Revista Argentina de Neuropsiciologia*, *11*, 1–8.

Cuetos, F., Martínez-García, C., & Suárez-Coalla, P. (2018). Prosodic Perception Problems in Spanish Dyslexia. *Scientific Studies of Reading*, *22*(1), 41–54. https://doi.org/10.1080/10888438.2017.1359273

Defior, S., Serrano, F., & Marín-Cano, M. J. (2008). El poder predictivo de las habilidades de conciencia fonológica en la lectura y escritura en castellano. En. *Estudios de Desarrollo Del Lenguaje y Educación. Studies on Language Development and Education*, (January), 339–347.

Duncan, L. G., Castro, S. L., Defior, S., Seymour, P. H. K., Baillie, S., Leybaert, J., … Serrano, F. (2013). Phonological development in relation to native language and literacy: Variations on a theme in six alphabetic orthographies. *Cognition*, *127*(3), 398–419. https://doi.org/10.1016/j.cognition.2013.02.009

Guardia Gutiérrez, P. A. (2003). Relaciones entre habilidades de alfabetización emergente y la lectura, desde nivel transición mayor a primero básico. *Psykhe*.

Guardia Gutiérrez, P. A. (2010). *The Effect of Linguistic , Phonetic and Lexical Factors on Phonological skills and Reading Acquisition in Spanish : A Longitudinal Study*. Cambridge.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*. https://doi.org/10.1007/BF02289447

Jiménez, J. E., & Ortiz, M. R. (1995). Prueba de segmentación lingüística (PSL). *Conciencia Fonológica y Aprendizaje de La Lectura: Teoría, Evaluación e Intervención. Madrid: Síntesis*.

Kam, C. C. S., & Meyer, J. P. (2015). How Careless Responding and Acquiescence Response Bias Can Influence Construct Dimensionality: The Case of Job Satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of Comprehensive Professional Development for Teachers of At-Risk Preschoolers. *Journal of Educational Psychology*, *101*(2), 448–465. https://doi.org/10.1037/a0013842

Lonigan, C. J., & Farver, J. M. (2002). Spanish preschool comprehensive test of phonological and print processing. *Unpublished Test, Florida State University*.

Manrique, A. M. B. de, & Graminga, S. (1984). La segmentación fonológica y silábica en niños de preescolar y primer grado. *Lectura y Vida*, *2*(analisis de la relacion para percibir el habla y leer), 12.

Manrique, A. M. B. de, & Signorini, A. (1994). Phonological awareness, spelling and reading abilities in Spanish-speaking children. *British Journal of Educational Psychology*, *64*(3), 429–439. https://doi.org/10.1111/j.2044-8279.1994.tb01114.x

Meira, Â., Cadime, I., & Leopoldina Viana, F. (2019). The structure of phonological awareness in European Portuguese: A study of preschool children. *Journal of Educational Research*, *112*(3), 367–376. https://doi.org/10.1080/00220671.2018.1530966

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., … Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, *29*, 65–77. https://doi.org/10.1016/j.learninstruc.2013.09.003

Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1998). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology*, *71*(1), 3–27. https://doi.org/10.1006/jecp.1998.2453

Papadopoulos, T. C., Spanoudis, G., & Kendeou, P. (2009). The Dimensionality of Phonological Abilities in Greek. *Reading Research Quarterly*, *44*(2), 127–143. https://doi.org/10.1598/rrq.44.2.2

Pearson, M. R. (2012). *Perfil de habilidades de lectura en español e inglés en una población argentina que asiste a escuelas bilingües*. Universidad Católica Argentina, Facultad de Psicología y Psicopedagogía.

Revelle, W. (2021). psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois.

Riccio, C. A., Amado, A., Sandra, J., Jane, H., Imhoff, B., & Denton, C. (2001). Cross-linguistic transfer of phonological processing: Development of a measure of phonological processing in spanish. *Bilingual Research Journal*, *25*(4), 583–603. https://doi.org/10.1080/15235882.2001.11074468

Runge, T. J., & Watkins, M. W. (2006). The structure of phonological awareness among kindergarten students. *School Psychology Review*, *35*(3), 370–386.

Vloedgraven, J. M. T., & Verhoeven, L. (2007). Screening of phonological awareness in the early elementary grades: An IRT approach. *Annals of Dyslexia*, *57*(1), 33–50. https://doi.org/10.1007/s11881-007-0001-2

Wackerle-Hollman, A., Durán, L., Brunner, S., Palma, J., Kohlmeier, T., & Rodriguez, M. C. (2019). Developing a Measure of Spanish Phonological Awareness for Preschool Age Children: Spanish Individual Growth and Development Indicators. *Educational Assessment*, *24*(1), 33–56. https://doi.org/10.1080/10627197.2018.1545570

Appendix B Study Two: supporting information

Table B. 1. Descriptive statistics for G1 measures

measure	mean	sd	min	max	skewness	kurtosis
IQ	15.17	5.55	0.00	33.00	-0.17	0.20
Vocabulary	0.83	0.12	0.27	1.00	-1.66	4.18
non-verbal STM	4.68	1.04	1.00	8.00	-0.08	1.28
verbal STM	4.39	0.81	2.00	6.00	-0.44	0.41
blending phonemes	0.53	0.23	0.06	1.00	0.12	-0.91
blending syllables	0.92	0.10	0.28	1.00	-2.13	6.89
onset matching phonemes	0.73	0.18	0.34	1.00	-0.19	-1.23
onset matching syllables	0.77	0.18	0.19	1.00	-0.62	-0.63
rhyme pseuodwords	0.65	0.20	0.20	1.00	0.06	-0.76
rhyme words	0.69	0.23	0.10	1.00	-0.15	-0.99
segmentation phonemes	0.51	0.28	0.04	1.00	0.26	-1.34
segmentation syllables	0.67	0.26	0.09	1.00	-0.23	-1.36
RAN colours	42.12	14.01	19.15	88.86	1.39	2.09
RAN objects	36.80	8.88	19.45	67.56	1.20	1.76
letters name	0.90	0.14	0.14	1.00	-2.38	6.04
letters sound	0.86	0.15	0.14	1.00	-2.14	4.78
accuracy words	0.75	0.34	0.00	1.00	-1.32	0.22
accuracy pseudowords	0.68	0.32	0.00	1.00	-1.14	-0.09
fluency (wpm)	21.13	15.98	0.00	99.00	1.13	2.38

Table B. 2. Model coefficients for the full model for decoding and fluency

Outcome: decoding

Fixed Effects						
	Estimate	Std. Error	t value	Pr(>|t|)	CI 95%	
Intercept	0.06	0.09	0.64	0.52	-0.13	0.25
Age	-0.03	0.05	-0.62	0.54	-0.13	0.07
Gender	0.01	0.10	0.09	0.93	-0.19	0.21
SES (L)	0.25	0.09	2.74	0.01	0.07	0.44
SES (Q)	0.12	0.11	1.16	0.25	-0.09	0.33
IQ	-0.03	0.06	-0.52	0.61	-0.14	0.08
Voc	0.08	0.06	1.45	0.15	-0.03	0.19
vSTM	0.15	0.06	2.50	0.01	0.03	0.26
nvSTM	0.15	0.05	2.78	0.01	0.04	0.26
RAN	-0.21	0.06	-3.83	0.00	-0.32	-0.10
LK	0.22	0.06	3.78	0.00	0.11	0.34
PA	0.09	0.05	1.65	0.10	-0.02	0.20

Random Effects		
	Variance	S.D.
School (Intercept)	0.06	0.25

Model fit		
R2	Marginal	Conditional
	0.39	0.46

Key: p-values for fixed effects calculated using Satterthwaites approximations. Confidence Intervals have been calculated using the Wald method.

Model equation: decoding ~ Age + gender + ses + IQ + VOC + nvSTM + vSTM + RAN + LK + PA + (1 | School)

outcome: fluency

Fixed Effects

	Estimate	Std. Error	t value	Pr(>|t|)	CI 95%	
Intercept	-0.14	0.10	-1.30	0.20	-0.35	0.07
Age	0.00	0.06	-0.03	0.97	-0.11	0.11
Gender	0.29	0.11	2.60	0.01	0.07	0.51
SES (L)	0.23	0.10	2.26	0.02	0.03	0.43
SES (Q)	0.13	0.11	1.13	0.26	-0.10	0.36
IQ	-0.01	0.06	-0.23	0.82	-0.13	0.10
Voc	0.07	0.06	1.11	0.27	-0.05	0.19
vSTM	0.10	0.06	1.60	0.11	-0.02	0.23
nvSTM	0.04	0.06	0.69	0.49	-0.07	0.15
RAN	-0.22	0.06	-3.55	0.00	-0.34	-0.10
LK	0.26	0.06	4.13	0.00	0.14	0.39
PA	0.03	0.06	0.58	0.56	-0.08	0.15

Random Effects

	Variance	S.D.
School (Intercept)	0.08	0.29

Model fit

R2	Marginal	Conditional
	0.30	0.38

Note: Models fitted though maximum likelihood, p-values for fixed effects calculated using Satterthwaites approximations. Confidence Intervals have been calculated using the Wald method.

Model equation: fluency ~ Age + gender + ses + IQ + VOC + nvSTM + vSTM + RAN + LK + PA + (1 | School)

Table B. 3. Details and comparison of nested models for decoding and fluency

Outcome: decoding

Sampling Units		N Subjects = 243 N Schools = 24								
Model specification	Model name	Nested / simpler Model	Fixed Effects added	Random Effects	Model fit				LRT Test against nested	
				School	AIC	BIC	LL	R2	df	X2
RE only	null	-	-	intercept	683.902	694.381	-338.951	0.125		
FE main effects	preliteracy	null	lk + ran + pa	intercept	601.498	622.456	-294.749	0.386	3	88.404
FE main effects	full	preliteracy	iq + age + gender + ses + voc + vstm + nvstm	intercept	586.672	635.574	-279.336	0.456	8	30.826

Outcome: fluency

Sampling Units		N Subjects = 243 N Schools = 24								
Model specification	Model name	Nested / simpler Model	Fixed Effects added	Random Effects	Model fit				LRT Test against nested	
				School	AIC	BIC	LL	R2	df	X2

RE only	null	-	-	intercept	672.554	682.983	-333.277	0.108		
FE main effects	preliteracy	null	lk + ran + pa	intercept	616.905	637.764	-302.453	0.313	3	61.649
FE main effects	full	preliteracy	iq + age + gender + ses + voc + vstm + nvstm	intercept	610.814	659.484	-291.407	0.380	8	22.091

Table B. 4. Details and comparison of nested models for decoding including interaction term

Outcome: decoding

Sampling Units		N Subjects = 243 N Schools = 24								
Model specification	Model name	Nested / simpler Model	Fixed Effects added	Random Effects	Model fit				LRT Test against nested	
				School	AIC	BIC	LL	R2	df	X2
FE main effects	full	preliteracy	age + gender + IQ + ses + voc + vstm + nvstm + lk + ran + pa	intercept	586.672	635.574	-279.336	0.456		
FE interaction	interaction	full	Ses:PA	intercept	576.98	632.87	-272.49	0.487	2	13.69

Model equation: decoding ~ Age + gender + ses + IQ + VOC + nvSTM + vSTM + RAN + LK + PA + SES:PA + (1 | School)

June 2021

Maldonado, Uruguay