Mind the orthography: revisiting the contribution of pre-reading phonological awareness to reading acquisition

Camila Zugarramurdi[1,2], Lucía Fernández[1,3], Marie Lallier[2], Juan Carlos Valle-Lisboa[1,3], Manuel Carreiras[2,4,5]

[1] Instituto de Fundamentos y Métodos en Psicología y Centro de Investigación Básica en Psicología (CIBPsi), Facultad de Psicología, Universidad de la República, Montevideo, Uruguay

[2] Basque Center on Cognition Brain and Language (BCBL), Donostia, Spain

[3] Centro Interdisciplinario en Cognición para la Enseñanza y el Aprendizaje (CICEA), Universidad de la República, Montevideo, Uruguay

[4]Ikerbasque, Basque Foundation for Science, Bilbao, Spain

[5]Departamento de Lengua Vasca y Comunicación, University of the Basque Country, Bilbao, Spain

Author note

Camila Zugarramurdi https://orcid.org/0000-0003-1675-0722
Marie Lallier https://orcid.org/0000-0003-4340-1296
Juan Valle-Lisboa https://orcid.org/0000-0002-5662-5937
Manuel Carreiras https://orcid.org/0000-0001-6726-7613

Juan Valle Lisboa is now at Sección Biofísica y Biología de Sistemas, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

Disclosures: The authors declare that the research was conducted is the absence of any commercial or financial relationships that should be construed as a potential conflict of interest. Manuel Carreiras and Juan Valle-Lisboa should be both considered senior authors of this publication.

Acknowledgments: This project was funded by ANII FSED_2_2015_1_120741 and ANII FSED_2_2016_1_131230 grants. Camila Zugarramurdi received a PhD Scholarship from Fundación Carolina. We thank Alvaro Cabana for helpful insights on data analysis, and all the families and schools involved in the study.

Data availability statement: The data that support the findings of this study are available from Camila Zugarramurdi (czugarramurdi@psico.edu.uy) upon reasonable request. This study's design and its analysis were not pre-registered.

Correspondence concerning this manuscript should be addressed to Juan Valle-Lisboa, Facultad de Ciencias, Iguá 4225, Montevideo 11600, Uruguay. Email: juancvl@fcien.edu.uy

Mind the orthography: revisiting the contribution of pre-reading phonological awareness to reading acquisition

Abstract

Reading acquisition is based on a set of preliteracy skills that lay the foundation for future reading abilities. Phonological awareness —the ability to identify and manipulate the sound units of oral language— has been reported to play a central role in reading acquisition. However, current evidence is mixed with respect to its universal contribution to reading acquisition across orthographies. This longitudinal study examines the development and contribution of phonological awareness to early reading skills in Spanish, a transparent orthography. The results of a comprehensive battery of phonological awareness skills in a large sample of children (time 1 n = 616, 296 females, mean age 5.6, from middle to high socioeconomic backgrounds; time 2 n = 397) with no reading experience at study onset suggest that the development of phonological awareness is delayed in Spanish. Furthermore, our results show that phonological awareness does not contribute to the prediction of reading acquisition above and beyond other preliteracy skills. Letter knowledge indexes children's ability to identify phonemes and thus takes a more central role in the prediction of early reading skills. Therefore, we underscore the need to thoughtfully address the distinctive features of the reading acquisition process across orthographies, which should be taken into account in models of reading and learning to read.

Keywords:

Phonological awareness, decoding, reading, Spanish, transparent orthography, longitudinal

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, Wouters, van Wieringen, & Ghesquière, 2007; Lyytinen et al., 2006; Muter, Hulme, Snowling, & Stevenson, 2004; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). Phonological awareness (PA) is a central construct in most, if not all, models of reading acquisition (for a meta-analysis see Melby-Lervåg, Lyster, & Hulme, 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, Fletcher, Snowling, & Scanlon, 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, Wagner, & Rashotte, 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.

Current issues on the PA-reading relation

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, Marinus, Coltheart, & Castles, 2015). Orthographic consistency, in turn, modulates the developmental trajectories of reading acquisition: high decoding levels are achieved faster in more transparent orthographies (Seymour, Aro, & Erskine, 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., 2012; 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, Petscher, Foorman, & Zhou, 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; Hulme, Snowling, Caravolas, & Carroll, 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, Spanoudis, & Kendeou, 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, Lonigan, Driscoll, Phillips, & Burgess, 2003). Different studies use a wide variety of different tasks, which makes it to compare them. 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, Majerus, & Poncelet, 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, Elwér, Samuelsson, Olson, & Byrne, 2019; Puolakanaho et al., 2007; Vaessen et al., 2010), others have not (De Jong & Van der Leij, 2003; Defior, Serrano, & Marín-Cano, 2008; Georgiou, Torppa, Manolitsis, Lyytinen, & Parrila, 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.

Additionally, 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., 2007). 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, Torppa, & Uppstad, 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.

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.

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, which 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 in more transparent orthographies 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.

Methods

Participants and procedure

Data was collected longitudinally in two consecutive school years starting with the last year of Kindergarten (K5), followed by 1st grade (G1). All children were native Spanish speakers. Children were recruited from 26 public schools in Montevideo, Uruguay, with middle to high socio-economic status, according to the National School Administration classification. Schools were either part-time or full-time. All children attending K5 level at time 1 were invited to take part in the study. Only those whose parents signed a consent form, in accordance with the Declaration of Helsinki, took part (N = 616, 296 females, M age = 5.6 years).

Sample size was estimated based on expected dropout, prevalence of reading difficulties, and power calculations. Power calculations using GPower3.1 (Faul, Erdfelder, Buchner & Lang, 2009) show that, for a multiple regression with 11 factors (see below), and an expected effect size $f^2 = 0.1$, a sample of at least 262 children is needed to obtain a power of .95 at .05 level. Dropout was estimated at 30%, given the study was conducted in the school setting, longitudinally; and prevalence of reading difficulties was estimated at 10%. Thus, an initial sample of approximately 600 children was targeted for.

The final sample, including children that completed both stages of data collection, consisted of 397 children (181 females, M age = 5.6). Dropout was mainly due to children that switched schools between time points. Nine children were additionally excluded because they did not complete at least half of the tasks. Dropout analysis showed significant differences in SES ($X^2(2, N = 616) = 6.82$, $p = .033$), with those dropping out showing a larger proportion of children from low SES. No other variable of interest showed any significant differences between groups (Age: $t(614)= 0.55$, $p = .58$; Gender: $X^2(1, N = 616) = .70$, $p = .40$, IQ: $t(614)= 1.29$, $p = .20$, letter knowledge: $t(614)= 1.80$, $p = .07$, phonological awareness: $t(614)= 0.32$, $p = .75$, and rapid automatized naming: $t(614)= -1.15$, $p = .25$)

Whenever justified, missing data was imputed among independent variables (i.e., excluding reading). The only exception being SES, since there are not enough social variables to make a valid imputation (see below). Demographic information (age, gender, and maternal education as a proxy for socioeconomic status) was obtained from a national database from the educational system (ANEP). Seventy-five children were excluded due to missing SES data. Missing data was present across measured variables (M = 14%, min = 10%, max = 18%) and was handled through random forest imputation via the *missforest* package in R (Eckert, Vaden, & Gebregziabher, 2018; Stekhoven

& Buehlmann, 2012). This method predicts the missing data taking into account all the other data and t has been shown to be highly effective even when there is low or moderate correlation between variables (Tang & Ishwaran, 2017). No dependent variables (i.e., reading) were included in the imputed procedure. Normalized root mean square error of approximation for the imputation of continuous variables was 0.30 (Stekhoven & Buehlmann, 2012).

Assessment was carried out at the schools, in groups of four to five children, except for the reading and RAN tasks which were assessed individually. Sessions lasted between 10 and 15 minutes. All tasks were presented digitally through individual tablets, except for reading and RAN. In order to do so, we created an Android-based application in videogame format, called *Lexiland* (Zugarramurdi, Fernández, Lallier, Carreiras, Valle-Lisboa, in press). Instructions and stimuli were pre-recorded and delivered though headphones. All tasks began with two to three examples, followed by two to three practice trials with feedback. Children were allowed to repeat practice trials freely. Two research assistants monitored task performance and were available to clarify instructions on demand.

All the procedures were approved by the Ethics Committee of the School of Psychology at Universidad de la República; date of approval February 17th, 2016, under the study title: “Design of a digital assessment battery to detect reading difficulties”.

Measures

Phonological awareness. Phonological awareness was assessed by 4 different tasks: segmentation, blending, onset matching, and rhyming. In *Segmentation* (22 syllabic items, 28 phonemic items) children heard a word —with their corresponding image to reduce memory load— and had to segment it. In order to avoid verbal responses, 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. This task is similar to the task called tapping by other authors (see Bryant, MacLean, Bradley &

Crossland, 1990), a name we avoid since *Lexiland* includes a different tapping task (tapping to a rhythm). In the *Blending* test (18 syllabic items, 16 phonemic items), children heard a sequence of syllables or phonemes and had to blend them into a word. The phonemes were sounded following usual procedures in phonics instruction (McGuinness, 2004). Children gave their response by pressing the image corresponding to the blended word on screen. The target image was presented along with three distractors (one semantically related, one phonologically related, one unrelated). The location of the correct response was randomized across trials. Within each grain size, items ranged from two to four syllables, and four to six phonemes. In the *Onset matching* test (27 syllabic items, 32 phonemic items) children heard pairs of words —with their corresponding images— and had to indicate whether both words started with the same phoneme/syllable or not. Children gave their response by pressing on a tick or a cross button. Phonemic items with matching onsets shared the first phoneme for CV word onsets and the first two phonemes for CCV word onset (*flaco-flecha*). Items with non-matching onsets did not share any phonemes in the first syllable. Half of the items had matching onsets and half had non-matching onsets. In the *Rhyming* test (10-word items, 10-pseudowords items), children heard two words —with its corresponding images— or two pseudowords and had to indicate whether they rhymed or not. Children gave their response by pressing on a tick or a cross button. Half of the items rhymed, and half did not. For each task, the score was calculated as mean accuracy.

We used Item response theory (IRT) to estimate each subject's latent phonological awareness score from observed responses. We used a 2-parameter model where the hit probability on each item is modeled as a logistic curve defined by two item parameters: discrimination and difficulty. The model allowed us to estimate the discrimination and difficulty of each item and the subject latent ability. 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). In the IRT context, *information* is used to replace the concept of reliability. Information is conceptualized as a function of model parameters and varies with the level of ability. The information of the test is defined as the average information of its items.

Letter knowledge (LK). Letter knowledge was assessed separately for letter sounds and names. In each subtask, children heard the name/sound of a letter and had to choose the answer by tapping the screen to select one of three options: the target letter, a visually similar distractor (Boles & Clifford, 1989), or an unrelated distractor. There were 22 items of each type (for a total of 44). For each task, the score was calculated as mean accuracy.

Rapid Automatized naming (RAN). Children were presented with a 6 x 5 array of five items each repeated six times and were asked to name them as quickly and as accurately as possible. Items were either objects (*gato, jugo, mano, silla, queso* [cat, juice, hand, chair, cheese, respectively]) or colours (*azul, negro, rojo, verde, blanco* [blue, black, red, green, white]). All subtasks were preceded by a familiarization phase. The score was calculated as total response time.

Vocabulary (VOC). Receptive vocabulary was measured through the noun subset of the BEST vocabulary test (De Bruin, Carreiras, & Duñabeitia, 2017). Children heard a word and had to select one out of four images.

Short-term memory. Verbal short-term memory (vSTM) was assessed through an adaptation of the task described in Martinez Peres, Majerus and Poncelet (2012). Children heard a sequence of monosyllabic words, followed by their corresponding images on the screen (*sol, pan, tren, rey, flor, pez* [sun, bread, train, king, flower, fish]). Children were asked to order the images according to the order of presentation of the words heard. The sequence ranged from two to six items; the test included three trials of each sequence length. The order of presentation in each trial was randomized. The score was the maximum number of items recalled. Non-verbal short-term memory (nvSTM) was assessed through an adaptation of the Corsi Block task (Corsi, 1972). Blocks were replaced by pictures of pigs to make the task more attractive to children. Sequences ranged from two to eight items, four trials per sequence length. Testing was interrupted if three errors were made on four consecutive trials of the same length. The score was the maximum number of items recalled.

Nonverbal IQ. Nonverbal IQ was measured using the Matrix Reasoning subtest of the Spanish version of the WPPSI (Wechsler, 2001). Scores were computed as the maximum level achieved, following the WPPSI scoring system.

Phonological Awareness, Letter knowledge, RAN, Vocabulary, Short-term memory, and nonverbal IQ measures reported were used in both K5 and G1.

Reading.

At Time 1 (K5), a list of 15 words and 15 pseudowords was presented on paper and children were asked to read them aloud. All items consisted of 2-syllable high frequency words with varying syllabic complexity including CV and CCV onset items. Pseudowords were constructed from the list of words using Wuggy (Keuleers & Brysbaert, 2010).

At time 2 (G1), reading assessment included two tasks: decoding and fluency. *Decoding*. A list of 30 words and 30 pseudowords was presented digitally, one word per screen. Items consisted of two to three syllable words of medium frequency with simple and complex syllables. Mean accuracy was computed for each child. *Fluency*. A two-minute reading test. It consisted in reading as quickly and as accurately as possible a text of 278 words on paper, within a maximum of 2 minutes. This was a Spanish adaptation of the Alouette French test (Lefavrais, 2005). The number of correct words read per minute was computed for each child.

Data and materials are available upon request to the first author. This study’s design and its analysis were not pre-registered.

Results

The rationale for the analysis was as follows. First, we studied the development of phonological awareness from K5 to G1, at the syllabic and phonemic level. In order to control for measurement error, we computed latent ability scores of phonological awareness for each child. Next, we assessed individual reading levels in K5 to identify children who could read and children who could not read at all, and compared the 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. All analyses were performed using R software (R Core Team, 2018).

Table 1

Descriptive statistics for K5 measures and G1 reading

time		M	SD	min	max	skewness	kurtosis	reliability	chance
K5	Age	5.60	0.29	5.10	6.20	0.03	-1.21	-	-
K5	IQ	10.09	5.42	0.00	28.00	0.60	0.08	0.90	-
K5	Vocabulary	0.83	0.12	0.27	1.00	-1.66	4.18	0.87	0.25
K5	non-verbal STM	3.57	1.23	1.00	6.00	-0.26	-0.53	0.77	-
K5	verbal STM	3.62	1.04	1.00	6.00	-0.05	-0.44	0.75	-
K5	blending phonemes	0.31	0.18	0.00	0.94	1.36	1.87	0.64	0.38
K5	blending syllables	0.83	0.15	0.21	1.00	-1.46	2.37	0.79	0.37
K5	onset matching phonemes	0.54	0.12	0.31	0.97	1.46	2.19	0.64	0.50
K5	onset matching syllables	0.59	0.15	0.19	1.00	0.81	-0.04	0.74	0.50
K5	rhyme pseudowords	0.54	0.16	0.00	1.00	0.34	0.99	0.41	0.50
K5	rhyme words	0.57	0.17	0.10	1.00	0.57	0.53	0.44	0.50
K5	segmentation phonemes	0.27	0.12	0.07	0.96	2.00	7.28	0.62	0.25
K5	segmentation syllables	0.41	0.17	0.09	1.00	1.36	2.13	0.76	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	0.87	0.33
K5	letter sound	0.55	0.21	0.14	1.00	0.32	-0.85	0.81	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 McDondald’s omega (total). Reliability of the average Phonological Awareness score is 0.85 (when including all items) and for Letter Knowledge is 0.91. N = 388.

Descriptive statistics for K5 measures and G1 reading are reported in Table 1 (See Table S1 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, 95% CI [0.49 – 0.63], *p* < .001; LK r = 0.77, [0.73 – 0.81], *p* < .001; decoding r = 0.96, [0.95 – 0.96], *p* < .001). For PA, latent ability scores were computed from all tasks (See section on IRT analysis). Correlations among all variables measured in K5 and reading measured in G1 were studied to assess collinearity issues for model building (Table 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 2

Pearson correlation coefficients for K5 variables and G1 reading measures

	time		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 < .001, ** p < .01 false discovery rate correction

Note. PA: phonological awareness, Voc: vocabulary, nvSTM: non-verbal short-term memory, vSTM: verbal short-term memory, RAN: rapid automatized naming, LK: letter knowledge. N=388.

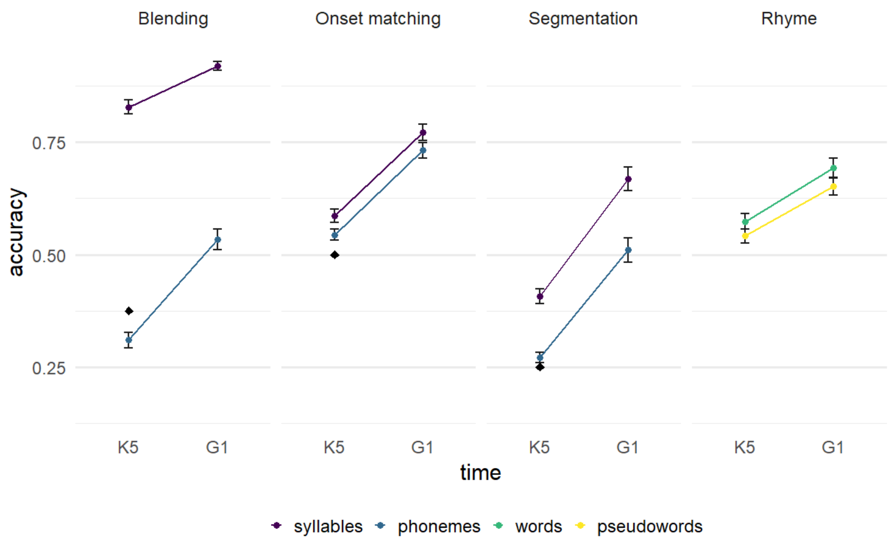
Phonemic Awareness shows Floor Effects in Kindergarten

In order to evaluate performance across measures and time points (Figure 1), we first performed one-sample t-tests of accuracy scores against chance, since all tasks were presented in a multiple-choice format. Children performed better than chance across all tasks (p < .001), except for blending phonemes, where average performance was significantly below chance level (M= 0.31, chance = 0.37, 95% CI [0.29, 0.33], t = -7.038, df = 387, *p* < .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 [0.53, 0.55]).

Next, since PA skills were assessed both in K5 and G1, we could evaluate growth in PA skills across time (Figure 1). A linear mixed effect model with 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 1

Phonological Awareness Performance Across Tasks, Grain Size and Time Points



Note. 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. N = 388.

IRT Analysis shows Progression of Difficulty from Syllabic to Phonemic items

Since the PA construct was measured by four tasks varying in difficulty and cognitive load, we estimated a latent ability score for each child by combining all tasks measured in K5. This served two ends. First, it enabled us to directly compare difficulty levels among tasks. Second, it controlled 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, 2009). Model fit indices showed excellent fit (additional details on model fit available in Supporting information). Twelve items were excluded due to extreme difficulty parameters. The final, reduced, model included 151 phonological awareness items.

Person-level analysis. Having established adequate model fit, latent ability scores were computed for each child from the reduced 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 .99. Obtained latent ability scores were normally distributed around 0 (min: -2.9, max: 2.7).

Item-level analysis. Difficulty was examined for each task and grain size (that is, syllabic vs. phonemic items). Overall average difficulty for the reduced model was .5. Tasks arranged from less to more difficult were blending (M = −1.23) < rhyme (M = 0.33) < onset matching (M= 0.61) < segmentation (M= 1.87). Pairwise comparisons through two-sample t-test showed significant differences between blending and onset matching (t(147) = −184, *p* = .019) and between blending and segmentation (t(147) = −3.10, *p* < .001). With respect to grain size, syllabic items were significantly less difficult than phonemic ones (M syllables = −0.62, M phonemes = 1.65, t(129) = 2.27, *p* < .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 (M = 0.24) < blending (M = 0.35) < rhyme (M = 0.38) < onset matching (M = 1.16). Pairwise comparisons showed significant differences between segmentation and onset matching (t(147) = 0.92, p < .001), blending and onset matching (t(147) = −0.80, p < .001), and rhyme and onset matching (t(147) = 0.77, p < .001). No significant differences were observed in discrimination parameters between syllabic and phonemic items.

Finally, we computed test information in order to assess internal consistency. We found that the peak of the information function is located at 0.03, suggesting the estimation of latent ability scores for the PA construct are most precise at average latent ability score levels, with a standard deviation of 0.19. Thus, error of measurement was very low for most children with scores around average, and increased towards the edges, as expected, with a standard deviation of approximately 0.4 for latent ability scores of 2 and -2. For approximately 95% of the sample, which falls between -2 and 2 standard deviations of the mean, the error in estimation was very low.

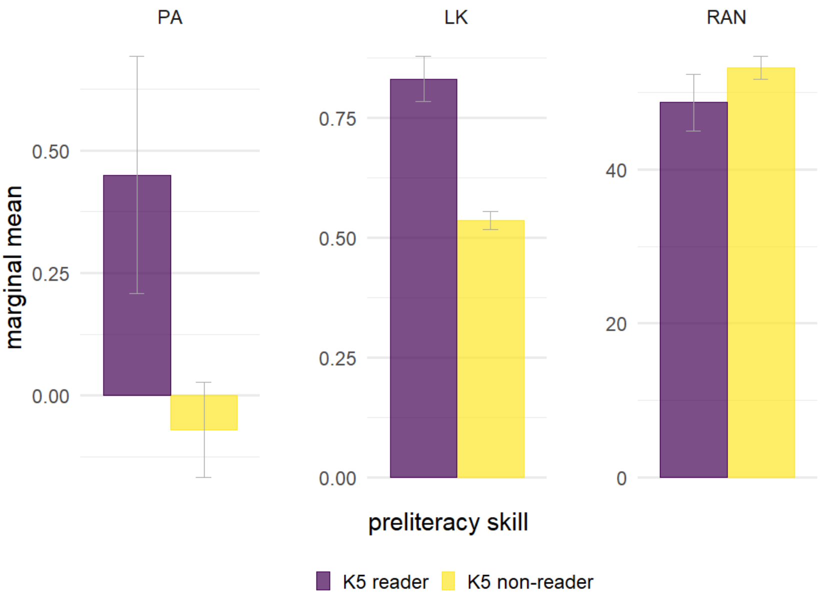
Preliteracy Skills differ between K5 Readers and Non-readers

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 make sure that our results refer to pre-readers avoiding reciprocal effects, we conservatively 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 because it excludes the use of any familiar whole-word recognition strategies, which do not imply the decoding ability.

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 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 2

Preliteracy Skills Performance of K5 Readers vs. Non-Readers at Time 1



Note. 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. N readers = 54, N non-readers = 334.

RAN and LK, but not PA, Uniquely Contribute to Reading Skills

In order to evaluate the unique contribution of PA to early reading abilities, that is, variance explained while controlling for relevant covariates, we run 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. The inclusion of the PA variable in the regression models presented a challenge due to a. the floor effects observed for phonemic items, and b. the low reliability for rhyme tasks. In order to overcome these challenges, we used latent ability scores obtained from the IRT analysis (See previous section about IRT Analysis), excluding the rhyme tasks. Additionally, all reported models were also fit with latent ability scores for phonemic awareness and syllabic awareness items separately. Results for the full model remained the same and are thus not reported.

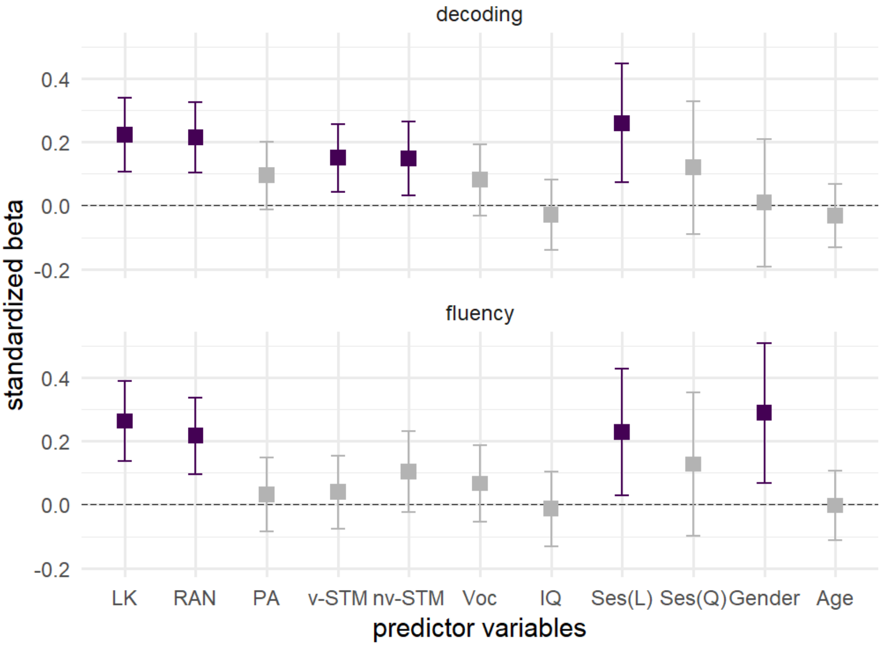
School was included as random intercept to account for the nesting of children across schools. Age, Gender, IQ, Vocabulary, vSTM, nvSTM, and Maternal Education, as a proxy for socioeconomic status (SES), were included as control variables. vSTM was treated as a control variable in order to focus on the 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 reported in Table S2.

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 3), which included all relevant covariates in addition to preliteracy skills, LK and RAN still contributed unique variance among preliteracy skills (Table S3). Crucially, PPA no longer contributed unique variance to decoding. Among covariates, vSTM, nvSTM, and SES all contributed unique variance to decoding. For fluency, SES 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 %.3.6 % to fluency. Both full models (accuracy and fluency) significantly improved model fit as compared to the preliteracy skills models.

Figure 3

Regression Coefficients for Prediction Model of Reading from Preliteracy Skills



Note. Prediction model coefficients for decoding (top panel, n = 243) and fluency (bottom panel, n = 239). 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). LK: Letter Knowledge, RAN: Rapid Automatized Naming, vSTM: verbal Short-Term Memory, nv-STM: non-verbal Short-Term Memory, Voc: Vocabulary, SES: Socio-economic Status. 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), non-verbal short-term memory (model 2), and SES (model 3). The only significant interaction effect observed was for PPA and SES (Table S4). 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 $X^2(2) = 13.69$, $p < .001$). In order to check the sensitivity of our results to changes in the cut-off value separating readers from non-readers, we re-run the analysis with a threshold of 5 pseudowords (Supporting information Table S5). The results did not change after this manipulation. Moreover, we evaluated a model of decoding (which is the dependent variable more likely to show PA effects) including K5 readers and a categorical variable representing whether they read or not at K5 (Table S6). The model does not show an independent effect of PA above and beyond all other predictors.

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. As in many previous studies, PA measured in K5 is correlated with reading scores measured in G1. Nevertheless, in two regression models of decoding and fluency, we showed that pre-reading phonological awareness (PPA) does not uniquely contribute to the prediction of early reading acquisition above and beyond other preliteracy skills, while controlling for several relevant covariates. We have also shown that our failure to find a contribution of PPA above and beyond the other factors is not due to lack of power. Instead, we showed that LK and RAN (and vSTM in the case of decoding) are the most relevant predictors of early reading skills. Crucially, our prediction models can account for large amounts of variance (38% and 45% for decoding and fluency respectively) 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.

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" (Defior et al., 2008; Mann & Wimmer, 2002). Moreover, it is possible that in opaque orthographies PA is stimulated during kindergarten and that this is not the case in transparent orthographies, as it is felt by teachers and curriculum designers that this is a skill that can be learned easily. In line with this possibility, Ziegler and Goswami (2005) review evidence that children learning to read in transparent orthographies show little phonemic awareness in K5 (e.g., Italian or Greek) but develop it quickly during first grade.

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 exclusion of K5 readers in our model could raise the question of whether we are artificially reducing the effect that PA can have in reading in G1. However, as shown above, results are robust even when changing the threshold separating readers from non-readers.

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. In addition, the reported pattern of results showing a significant correlation between PA skills in K5 and reading scores in G1, **the** growth of PA skills from K5 to G1, the increased performance for syllabic to phonemic items, the separation across cognitive operations, and the discrimination between K5 readers and non-readers, strongly suggest that the null contribution does not stem from random behaviour. 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 see 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 prediction. 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 to 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 transparent orthographies, casts doubt on a universal role for PPA as predictor of reading acquisition. Nevertheless, 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 co-requisite rather than as a prerequisite of reading acquisition. We believe that a dynamic interplay between PPA and LK can accommodate the observed pattern. Following Mann and 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 (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, Ellefson, Solity, & Chater, 2011). In less transparent orthographies, where the number of phonemes tends to be larger than the number of letters, thus requiring more complex graphemes to represent them, the ability to identify and manipulate phonemes (i.e., PA) has a larger explanatory value than knowing the letters. 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, Erskine, Hämäläinen, Torppa, & Ronimus, 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. What we measure as LK includes in part what we measure as PA.

In sum, the unique contributions of PA and LK as longitudinal predictors of decoding abilities are the result of a combination of kindergarten instructional practices, the home literacy environment and the differential information content contributed by LK and PA across orthographies. In a transparent orthography, a child with good LK, good memory, and good lexical access (i.e., RAN), only needs to learn to synthesize to be able to decode words, whereas this is not enough in an opaque orthography, and the child still needs to refine her knowledge of the grapheme – phoneme mappings, as it is not one-to-one. That is why in more transparent orthographies reading and PA develop concurrently (and those children are excluded from our prediction), whereas there is an intermediate step of good PA and letter knowledge of non-readers in more opaque orthographies.

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, 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, the response format used helps separate PA from speech production and memory. In this way, our PA tasks are tapping into the PA construct, albeit through a different measurement than usual.

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 (Wagner & Torgesen, 1989*;* Knoop-van Campen, Segers, & Verhoeven, 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.

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 acquired 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. As mentioned before, it might underlie the little development of PA skills in K5. 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 that in a transparent orthography, among children who had not yet developed reading, PPA does not make a unique contribution to the prediction of later reading acquisition, that is, when other variables such as letter knowledge, memory and RAN are included. 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 dynamic interplay 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.

References

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., … 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

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*. *Journal of Abnormal Child Psychology* (Vol. 33, pp. 205–217). https://doi.org/10.1007/s10802-005-1828-9

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

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

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

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., … 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

Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. Developmental Psychology, 26(3), 429-438. https://doi.org/10.1037/0012-1649.26.3.429

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. *Psychological Science*, *23*(6), 678–686. https://doi.org/10.1177/0956797611434536

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

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

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

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

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

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

Eckert, M. A., Vaden, K. I., & Gebregziabher, M. (2018). Reading profiles in multi-site data with missingness. *Frontiers in Psychology*, *9*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2018.00644

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

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

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

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

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., 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.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

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

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

Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G., … 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

Lefavrais, P. (2005). Test de l'Alouette Révisé. *Paris: Editions Du Centre de Psychologie Appliquée*.

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. (2006). 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

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

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

McGuinness, D. (2004). *Early reading instruction*. MIT Press.

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

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. 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

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

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

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

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. Vienna, Austria.

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

Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling. *Journal Of Statistical Software*, *17*(5), 1–25.

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

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

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

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

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

Stekhoven, D. J., & Buehlmann, P. (2012). MissForest - non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118.

Tang, F., & Ishwaran, H. (2017). Random Forest Missing Data Algorithms. *Statistical analysis and data mining*, *10*(6), 363–377. https://doi.org/10.1002/sam.11348

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

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

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.

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. (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

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. Psychological Bulletin, 101(2), 192-212. https://doi.org/10.1037/0033-2909.101.2.192

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

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

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., & 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

Zugarramurdi, C., Fernández, L., Lallier, M., Carreiras, M., & Valle-Lisboa, J. C. (in press). Lexiland: a Tablet-based Universal Screener for Reading Difficulties in the School Context. *Journal of Educational Computing Research.*