Original Research Article

Journal of Educational Computing Research
2022, Vol. 0(0) 1–28

DOI: 10.1177/07356331221074300
journals.sagepub.com/home/jec
SAGE

Lexiland: A Tablet-based Universal Screener for Reading Difficulties in the School Context

Camila Zugarramurdi[1,2,3], Lucía Fernández[1,3], Marie Lallier[2], Manuel Carreiras[2,4,5], and Juan C. Valle-Lisboa[1,3,6]

Abstract

Massive and timely screening of the student population for early signs of reading difficulties is needed to implement timely effective remediation of these difficulties. However, traditional approaches are costly and hard to apply. Here, we present Lexiland, a tablet-based reading assessment tool for kindergarten and primary school children developed to be applied in school settings with minimal personnel intervention. Following a story line, players help a character of the game perform several tasks that measure different predictors of reading outcomes. Most of the tasks that usually involve a verbal response were switched to receptive tasks to demand a touch-screen response only. The tablet application was administered to a sample of *N* = 616 5-yo kindergarten children and to a sub-sample of these children twice during the following two years (First and Second Grades). Applying logistic regression and cross-validation, we selected a reduced subset of tasks that can predict with great sensitivity

[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
[6]Sección Biofísica y Biología de Sistemas, Facultad de Ciencias, Universidad de la República, Uruguay

Corresponding Author:
Juan C. Valle-Lisboa, Sección Biofísica y Biología de Sistemas, Facultad de Ciencias, Universidad de la República. Iguá 4225, Montevideo 11600, Uruguay.
Email: juancvl@fcien.edu.uy

and specificity, whether a five-year-old child will have reading difficulties by the end of first grade (sensitivity 90% and specificity 76%) and two years later (sensitivity 90% and specificity 61%). Importantly, Lexiland is a scalable tool to implement universal screening, given the increasing availability of devices able to run android and iOS applications.

Keywords

reading, evaluation, elementary education, games, cognitive, development, assessment, digital

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 second grade. In third grade, if he or she is still struggling, then he or she 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 program. This is the current protocol in many countries (Seidenberg, 2017, Chapter 11). While the reasoning behind this sounds intuitive, 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 socio-emotional development, and higher rates of depression and anxiety (Arnold et al., 2005; Sprenger-Charolles et al., 2011). Moreover, poor readers accumulate less reading experience than their peers, thus acquiring less vocabulary, in a downwards spiral known as a "Mathew effect", where rich get richer and 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, prominently, phonological awareness (PA), letter knowledge (LK), and rapid automatized naming (RAN), among others. Briefly, 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. Letter knowledge 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, colors, letters, or numbers that the participant has to name as quickly and accurately as possible. Numerous laboratory studies have shown the predictive value of preliteracy skills in estimating future reading outcomes (Boets et al., 2007; Lyytinen et al., 2006; Muter et al., 2004; Schatschneider et al., 2004).

Thus, by assessing 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 countries as different as 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.

Desirable Features of a Universal Early Screener

A Universal Screener Needs to be Timely

Even though 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), extrapolating to universal applications 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 assessment tasks to be "gamified", increasing children's motivation and engagement, and making it possible for collective self-assessment (Hautala et al., 2020), thus eliminating the need for trained applicants. Second, responses can be recorded and automatically processed, also without the need for trained staff. Third, data collection is 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. For this promise to be fulfilled, the screener has to show validity and reliability values at least comparable to those of traditional tests.

A Universal Screener Needs to Have High Sensitivity and Specificity

As in all screening and binary classification methods, the quality of the method is expressed by two quantities, the *sensitivity* of the method refers to the proportion of positive cases (i.e., those belonging to a class) that the method classifies as positive while the *specificity* refers to the proportion of negative cases that the method correctly identifies as not belonging to the class. 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 0% specificity, as it would erroneously classify all children who are not at-risk as at-risk. Previous studies in reading have obtained the best sensitivity and specificity when behavioral 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 behavioral 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.

Digital and Game Based Screeners of Future Reading Skills

There are currently many commercially available screening tools for reading difficulties. A search in the Academic Screening Tools from the National Center on Intensive Intervention (https://charts.intensiveintervention.org/ascreening) shows ten screening tools targeted at elementary (K5) children: DIBELS, FastBridge, Imagine Learning, i-Ready Diagnostic, i STEEP, Lexia RAPID Assessment, MAP, mCLASS Reading 3D, PALS, and TPRI Early Reading Assessment. However, its technical rigor is highly variable, with some of them showing convincing evidence for their classification accuracy, some of them showing unconvincing evidence, and some showing no evidence at all. Only few of them report its validity and generalizability, and the ones that exist, are published in technical reports (Baker et al., 2007), something that is also true of other popular tools like ISTATION (Basaraba et al., 2018), or NWEA (2019). In addition, none of them fulfills the requirements of scalability for universal use, since they are mostly based on one-on-one assessments, and most frequently on paper format.

The use of digital devices like tablets, cellular phones, or computers to apply educational or psychological assessments has been sought since the advent of digital technologies. It cannot be said that this is a new field, but lately it has been maturing into a fully developed one (Bennett, 2015; 2018; Neumann et al., 2019; Neumann & Neumann, 2019). In particular, for the assessment of children's abilities, game-based approaches seem to be particularly well suited (Hautala et al., 2020). In the early days, most digital technologies were compared with traditional paper and pencil methods, with mixed results with respect to the compatibility between the new methods and the traditional ones (Bennett et al., 2008). The maturation of the field brought the possibility of building new assessment tools based on cognitive principles and comparing the measurements of these tests directly to expected learning or psychological outcomes.

When searching the academic literature for digital and game-based screeners of reading, we found scarce evidence for pre-reading assessment tools. Rauschenberger and coworkers (2019) review screeners for readers and pre-readers. Most of the studies they found use small sample sizes and do not attempt to predict reading outcomes longitudinally. Moreover, few published studies attempt to predict reading behavior from pre-reading measures. For example, Drigas & Politi-Georgiousi (2019) review available game-based screeners for dyslexia targeted to the preschool years, and enumerate seven available serious games for screening dyslexia across different countries. However, while they were designed based on the science of reading, similarly to commercially available ones, none of them were experimentally validated through a longitudinal study that can effectively assess their classification accuracy (see, for example, Gaggi et al., 2012; Rauschenberger et al., 2020). A few exceptions exist. Carson et al. (2014) devised a game-based approach to assess PA in classrooms, allowing them to predict from kindergarten (5 y.o.) to the end of first grade the reading scores of children with great sensitivity and specificity. When combined with first-grade mid-term re-evaluations, they reached levels of sensitivity above 94% and specificity of 90%. In a pilot study using a game-based approach, Puolakanaho and Latvala (2017) could predict from 6.7 years of age (pre-school in Finland) to the end of first grade, which children would be considered slow readers or fast readers, achieving an impressive capacity to predict learning outcomes, with a sensitivity of 95.7% and a specificity of 81.8%. Singleton et al. (2000) could predict 50% of variance in reading at age 8 by assessing auditory verbal short-term memory (STM) and PA at age 5. One possible objection to these studies is that they did not use any validation procedure to estimate out-of-sample prediction capability and that they used a relatively small sample size. Despite these criticisms, these studies clearly show the advantages of using digital game-based tools for assessment. In this sense, these tools make universal screening a definite possibility.

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.

Early Predictors of Future Reading Outcomes: Preliteracy Skills

As briefly outlined in the first section of the Introduction, preliteracy skills (PA, LK and RAN, among others) have shown high predictive value in estimating future reading outcomes. PA has been consistently shown to be reduced in dyslexic readers (Melby-Lervåg et al., 2012) as well as in illiterates (Huettig & Pickering, 2019; Morais et al., 1987), which has led to assign PA as a central skill for reading acquisition. What is the role of PA in reading acquisition? The first step of reading acquisition encompasses the *decoding* stage. During decoding, children take each letter in a written word, transform it into its corresponding sound, and blend these sounds together, in order to access the words phonology and meaning. In order for this process to be successful, children need to be aware of the fact that oral language, perceived as a continuous acoustic stream, is composed of individual discrete units (phonemes) that are combined to form words; and they need to be able to access these phonemes in order to map them to their corresponding grapheme/letter. These awareness and access are what we call PA, thus, PA is central to the decoding process. A second element for decoding to be successful is

LK, since children need to know the letters (their shape, name and sound) in order to map them to their corresponding phonemes (Foulin, 2005). The third preliteracy skill, RAN, takes particular relevance in a following reading stage, that of automatizing decoding to attain reading fluency (Norton & Wolf, 2012). Initially, decoding is slow and effortful; however, it needs to be automatized so that cognitive resources can be redirected to higher levels of reading, such as meaning construction. Rapid automatized naming reflects children's naming speed and lexical access (that is, accessing words' morphology, phonology, semantics, and syntax) as well as children's abilities involved in shifting the visuo-attentional focus from one word to the next when processing multi-word sequences. Apart from these, other constructs have also been consistently linked to reading outcomes. These entail, broadly, oral language skills other than PA, such as grammatical knowledge and vocabulary. Although they are typically more relevant in later stages of reading acquisition, especially during reading comprehension, they are also indirectly linked to decoding through the preliteracy skills stated above (Dickinson et al., 2010).

Several previous studies have provided convincing evidence of the strong link between preliteracy skills and reading acquisition (Boets et al., 2007; Lyytinen et al., 2006; Muter et al., 2004; Schatschneider et al., 2004). For example, Hulme and colleagues (Hulme et al., 2015) conducted a longitudinal study of typically developing children and children at risk of reading difficulties (due to family risk), before and after school entry. In a sample of 246 children, they evaluated PA, LK, RAN and other oral language skills such as articulation, sentence structure, and vocabulary. After school entry, they evaluated word-level reading and reading comprehension skills. Results showed that PA, grapheme-phoneme knowledge (which involved LK and letter writing) and RAN at 4.5 years of age significantly predicted early word reading, single-word reading and spelling at 5.5 years of age, which in turn predicted reading comprehension at age 8. Many others have shown similar patterns in English and other languages (Furnes & Samuelsson, 2010, 2011; Georgiou et al., 2008; Landerl et al., 2013; Landerl & Wimmer, 2008; Lonigan et al., 2000; Lyytinen et al., 2006; although see Zugarramurdi et al., 2022 for a discussion on the effects of orthography on prediction patterns). However, they have mostly been conducted within laboratories, limiting its generalizability to more ecological contexts.

The Present Study

Given the available evidence, we hypothesized that screening for reading difficulties by measuring preliteracy skills digitally, with children playing in parallel, in the school setting, is both feasible and cost-effective. These measures should be tested for internal consistency, external validity, and, specially, for predictive validity. Thus, if we base the design of a screening tool on the science of reading, we expect these criteria can be met by creating a digital app. If this is the case, the tool can be used to easily predict which children will show reading problems, and more importantly, it will allow us to intervene early on.

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. Crucially during the development of the app, we had to create minigames that measure the relevant reading predictors, but requiring minimal or no intervention from applicators or raters. In order to assess this App's predictive longitudinal validity, an initial sample of more than 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 assessed on preliteracy and broader cognitive skills and predicted as *poor readers* or *typical readers* based on their reading skills at the end of first grade. We show that, even when used in parallel, in groups, in the school setting, the Lexiland screener attained high classification accuracy for first and second grade reading skills.

Methods

Participants

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, 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. The remaining 6% could not be tracked (most of them due to a mismatch between their ID number in our database and the one in the educational system). At Time 2, one of the schools dropped out of the study for scheduling reasons (2.5% of children), and the remaining children's parents did not sign the consent to continue with the study (11,5%). 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 2 and Time 3; 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 4–5 children. 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.

Task and Measures

Lexiland General Design

The Lexiland video game (Figure 1) was developed to assess preliteracy and general cognitive skills, targeted at children in the last year of kindergarten (K5) and first grade (G1). The tasks included assessments for: PA, LK, RAN, VOC, (STM, verbal and non-verbal), IQ, and reading. We measure predictors and factors that allowed us to compare our data to the other studies already cited. In order to increase children's motivation and engagement in autonomous play, tasks were 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, to automate data collection and processing. Thus, verbal responses were replaced by multiple

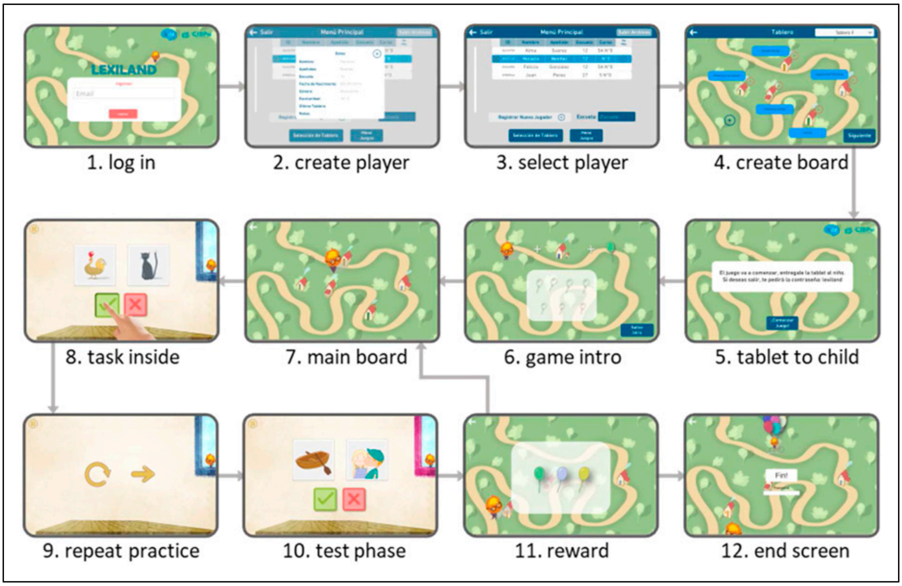

Figure 1. Lexiland® game flow. Screens 1 to 5 show the user interface for the adult user, Screens 6 to 12 show the user interface for children. Screens 8 to 10 are variable depending on the task being administered. The example depicts the Onset Matching task.

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. A comprehensive description of each task is available below, additional information on the Lexiland App design and tasks can be accessed in the supplementary material.

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. A word was presented aurally together with a picture of it and children were asked to segment it in either syllables or phonemes. In order to avoid verbal responses, together with the picture of the word, illustrations of dices corresponding to numbers two to four for syllables, and three to five for phonemes appeared on 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. The task consisted of 22 syllabic items and 28 phonemic items.

Blending. 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. The task consisted of 18 syllabic items and 16 phonemic items.

Onset Matching and Rhyme. 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 (Onset Matching) 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. The Onset Matching task consisted of 27 syllabic items and 32 phonemic items, the Rhyme task consisted of 10 word and 10 pseudoword items.

Letter Knowledge (LK)

Letter knowledge was assessed separately for letter name and letter sound. In each subtask, 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).

Rapid Automatized Naming (RAN)

Children were presented with an array of five items repeated six 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]), colors (*azul, negro, rojo, verde, blanco* [blue, black, red, green, white]), numbers (4, 5, 7, 8, 9) or capital letters (F, M, N, S, R). Notice that all items were disyllabic. Number of errors and total time were recorded. All children were presented with the four 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.

Vocabulary (VOC)

At Time 1, Receptive VOC 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 the Peabody Picture VOC Test (Dunn et al., 2006) was used. The procedure was the same on the tablet as it is on paper, except that the response was given by tapping on the screen.

Short-Term Memory (STM)

Verbal STM. Verbal STM was assessed through an adaptation of the task described in Martinez Perez et al. (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.

Non-Verbal STM. Visuo-spatial STM 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.

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).

Reading

Decoding (Time 1 and 2). We measured phonological decoding (Decoding, from now on) in two ways. 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. The 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.

Fluency (Time 2 and 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 (following Clark et al., 2021). The text was presented in paper. Number of words read and number of errors were recorded.

Comprehension (Time 2 and 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 of 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.

Data Analysis

All data analysis was performed using the R package (R Core Team, 2020)

Model Specification

Two logistic regression models were fit to the data in order to predict reader status. First, a *full model* with all cognitive and demographic variables was computed. Second, a *reduced model* was fit 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).

Cross Validation

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 reduce 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 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 labeled as positive (in this case, poor readers), otherwise it is labeled 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 common practice). Area under the curve values range from 0.5 to 1, where 0.5 indicates classification at chance level, and values above 0.8 are generally deemed as acceptable (Catts et al., 2009).

Table 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): $TN/FP + TN$

TP, FP, FN, TN refers to the total number of children in the corresponding condition; TPR: true positive rate; TNR: true negative rate.

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 *poor 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).

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 2. 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 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 2). On the other hand, 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).

Table 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.

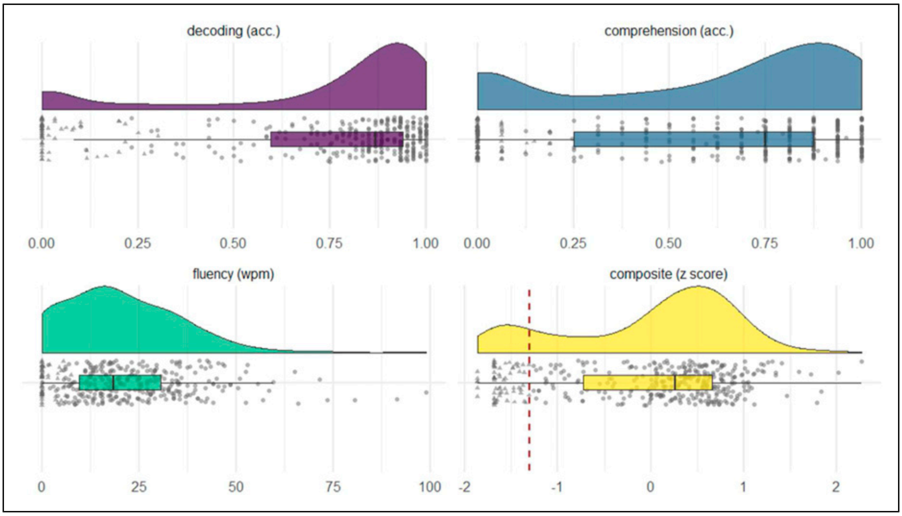

Figure 2. Distribution of reading scores for each reading measure. A bimodal distribution can be observed in all tasks. Triangles represent children that are classified as poor readers (PR), circles represent children classified as typical readers (TR). The dashed line indicates the cut-off threshold for reader status. Acc.: accuracy, wpm: words per minute.

Table 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.

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.

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, VOC, RAN, LK, and PA. Among these, only SES, non-verbal STM, LK, and PA were significant predictors of reader status above and beyond all other variables in the model (Table 3). The Nagelkerke pseudo R2 for this model was 70%.

With respect to the reduced model, which only included SES, non-verbal STM, LK, and PA, 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 reduces the number of variables that need to be measured—thus reducing assessment time—model fit is as good as the one of the full model.

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 analyzed: 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). 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.

Table 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. Bold rows show significant coefficients at the 95% level.

Cross Validation Performance for The Reduced Model

Finally, for the reduced model, cross-validation was performed to test its generalizability. Complete performance for each model iteration is presented in the ROC curve in Figure 4. 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.

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?

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 analyzed. 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

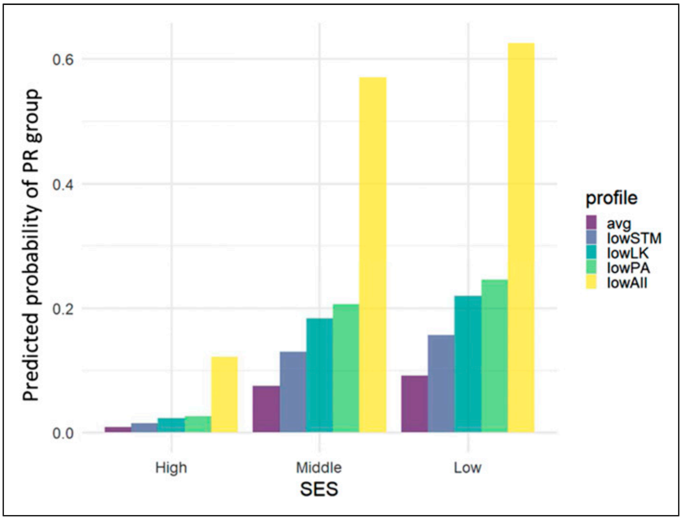

Figure 3. Predicted probabilities of belonging to the poor readers group by preliteracy skills profile and SES (reduced model).

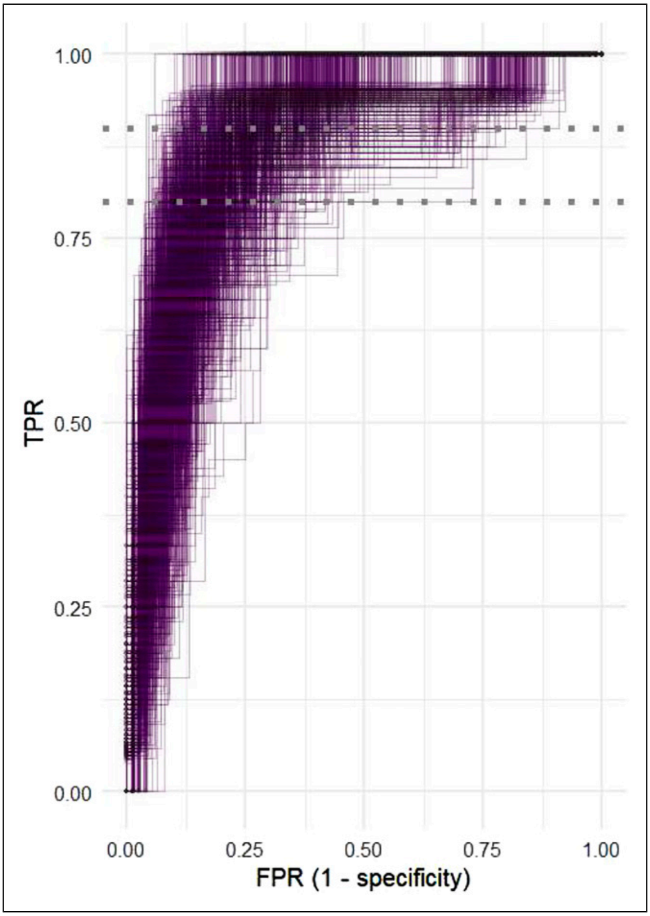

Figure 4. ROC curves for the reduced model. Gray dotted lines show 80% and 90% sensitivity (TPR) levels, corresponding to between 70% and 85% specificity levels. FPR: false positive rate, TPR: true positive rate.

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.

Stability in Model Prediction

The reduced model described in Table 3 was used to predict reader status in G2. This model was built using SES, non-verbal STM, LK, and PA 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.

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, LK, and PA), 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 maintaining 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) and it is in line with other studies that use digital media (Hautala et al., 2020; Puolakanaho & Latvala, 2017). Moreover, our analysis follows a robust validation procedure, which ensures that the sensitivity and specificity obtained will generalize outside of our sample. 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.

The ability to read is highly dependent on previous language knowledge which is highly variable between children of different socio-cultural environments (Hart & Risley, 2003). In line with this observation, socio-economic status was a significant predictor variable in all our models. Unfortunately, aiming for SES as a target for intervention is a much larger endeavor than focusing on preliteracy skills. Adequate teaching of letters names and/or, together with training in PA and STM —which can also be supported by digital Apps (Potier-Watkins & Dehaene, 2021; Richardson & Lyytinen, 2014)— 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)

Educational Implications

The present results show that Lexiland can be used as a universal screener. We are currently presenting these results to local authorities, while it is already being used in several studies in nearby countries (e. g., Argentina). Thus, we expect to be part of a National Policy for literacy. Lexiland can be used individually by teachers interested in having a more accurate 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.

Limitations

The present study was composed of an unselected sample of children attending K5 in middle and high SES public schools in Montevideo, Uruguay. 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 LK, PA, and non-verbal STM—in a representative sample of children with a broader range of SES statuses and from the entire country. It is important to note that the population of Montevideo represents approximately half of the country's population, so such an adjustment in sample is realistically within reach.

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 modeling 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.

Conclusions

To conclude, given the availability of digital devices such as tablets, phones and computers, early and timely identification of at-risk children is a feasible, inexpensive endeavor. Of particular importance, Lexiland's reduced version, which only entails three skills (PA, LK, and memory) measured in receptive tasks, attains high sensitivity and specificity, equivalent to the one in the full model. The fact that all tasks are measured receptively, without the need for any personnel, greatly increases the scalability of our tool, without sacrificing external and predictive validity.

It is well established that the earlier the identification, the more successful the intervention (Wanzek & Vaughn, 2007). Thus, the combination of early identification with effective teaching and remedial interventions (such as the program DALE, Diuk, 2019) can greatly improve the reading capabilities of at-risk children. Lexiland is a scalable solution that can be administered universally in varying conditions and settings.

Acknowledgments

We thank Alvaro Cabana for helpful insights on data analysis, and all the families and schools involved in the study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was funded by ANII FSED_2_2015_1_120741 and ANII FSED_2_2016_1_131230 grants to Juan Valle-Lisboa and Manuel Carreiras. Camila Zugarramurdi received a PhD Scholarship from Fundación Carolina.

ORCID iD

Juan C. Valle-Lisboa https://orcid.org/0000-0002-5662-5937

Supplemental Material

Supplemental material for this article is available online.

References

Andrade, O. V., 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.

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*, *33*(2), 205–217. https://doi.org/10.1007/s10802-005-1828-9.

Baker, D. L., Cummings, K. D., Good, R. H., & Smolkowski, K. (2007). Indicadores dinámicos del éxito en la lectura (IDEL): Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade (Technical Report No.1) Eugene, OR. https://dibels.uoregon.edu/sites/dibels1.uoregon.edu/files/IDEL_Instructional_Rec.pdf (accessed 17 September 2021).

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

Basaraba, D., Sparks, A., & Ketterlin-Geller, L. (2018). *Imagination station (Istation): Istation's indicators of progress (ISIP) Español validity study prekindergarten*. Southern Methodist University. Technical Report 18106.

Bennett, R. E. (2015). The changing nature of educational assessment. *Review of Research in Education*, *39*(1), 370–407. https://doi.org/10.3102/0091732X14554179.

Bennett, R. E. (2018). Educational assessment: What to watch in a rapidly changing world. *Educational Measurement: Issues and Practice*, *37*(4), 7–15. https://doi.org/10.1111/emip.12231.

Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *The Journal of Technology, Learning and Assessment*, *6*(9). Retrieved from https://ejournals.bc.edu/index.php/jtla/article/view/1639.

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.

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.

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.

Cardozo, S. (2016). Trayectorias educativas en la educación media PISA-L 2009-2014. INEEd - Grupo de estudios sobre Transiciones Educación-Trabajo (TET). https://www.ineed.edu.uy/images/pdf/trayectorias-educativas.pdf.

Carson, K., Boustead, T., & Gillon, G. (2014). Predicting reading outcomes in the classroom using a computer-based phonological awareness screening and monitoring assessment (Com-PASMA). *International Journal of Speech-Language Pathology*, *16*(6), 552–561. https://doi.org/10.3109/17549507.2013.855261.

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.

Clark, C., Guediche, S., & Lallier, M. (2021). Compensatory cross-modal effects of sentence context on visual word recognition in adults. *Reading and Writing*, ■■■(■■■), 1–19. https://doi.org/10.1007/s11145-021-10132-x.

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

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

De Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology*, *8*(April), 522. https://doi.org/10.3389/fpsyg.2017.00522.

Dickinson, D. K., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher*, *39*(4), 305–310. https://doi.org/10.3102/0013189X10370204.

Diuk, B. (2019). Those children who are left behind: A reading programme that works. In: *Improving early literacy outcomes* (pp. 155–169). Brill Sense.

Drigas, A. S., & Politi-Georgousi, S. (2019). ICTs as a distinct detection approach for dyslexia screening: A contemporary view. *International Journal of Online and Biomedical Engineering*, *15*(13), 46–60. https://doi.org/10.3991/ijoe.v15i13.11011.

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

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

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

Furnes, B., & Samuelsson, S. (2011). Phonological awareness and rapid automatized naming predicting early development in reading and spelling: Results from a cross-linguistic longitudinal study. *Learning and Individual Differences*, *21*(1), 85–95. https://doi.org/10.1016/j.lindif.2010.10.005.

Gaggi, O., Galiazzo, G., Palazzi, C., Facoetti, A., & Franceschini, S. (2012). A serious game for predicting the risk of developmental Dyslexia in pre-readers children. In 21st International Conference on Computer Communications and Networks (ICCCN), 2021, Munich, Germany, 1–5.

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology*, *100*(3), 566–580. https://doi.org/10.1037/0022-0663.100.3.566.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4–9. https://www.aft.org/sites/default/files/periodicals/TheEarlyCatastrophe.pdf.

Hautala, J., Heikkilä, R., Nieminen, L., Rantanen, V., Latvala, J.-M., & Richardson, U. (2020). Identification of reading difficulties by a digital game-based assessment technology. *Journal of Educational Computing Research*, *58*(5), 1003–1028. https://doi.org/10.1177/0735633120905309.

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. (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.

Huettig, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences*, *23*(6), 464–475. https://doi.org/10.1016/j.tics.2019.03.008.

Hulme, C., Nash, H. M., Gooch, D., Lervag, a., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of Dyslexia. *Psychological Science*, *26*(12), 1877–1886. https://doi.org/10.1177/0956797615603702.

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

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., O'Donovan, M., Williams, J., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Tõth, D., Honbolygõ, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Chaix, Y., Démonet, J. F., & Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(6), 686–694. https://doi.org/10.1111/jcpp.12029.

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*, *100*(1), 150–161. https://doi.org/10.1037/0022-0663.100.1.150.

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.

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.

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.

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.

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.

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.

Ministerio de Salud. (2017). *Detección precoz de hipoacusia congénita*. https://www.gub.uy/ministerio-salud-publica/sites/ministerio-salud-publica/files/documentos/publicaciones/Deteccion%20precoz%20de%20hipoacusia%20congenita.pdf.

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.

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.

Morais, J., Alegría, J., & Content, A. (1987). The relationships between segmental analysis and alphabetic literacy: An interactive view. *Cahiers de Psychologie Cognitive*, *7*(5), 415–438.

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.

Neumann, M. M., & Neumann, D. L. (2019). Validation of a touch screen tablet assessment of early literacy skills and a comparison with a traditional paper-based assessment. *International Journal of Research and Method in Education*, *42*(4), 385–398. https://doi.org/10.1080/1743727X.2018.1498078.

Neumann, M. M., Worrall, S., & Neumann, D. L. (2019). Validation of an expressive and receptive tablet assessment of early literacy. *Journal of Research on Technology in Education*, *51*(4), 326–341. https://doi.org/10.1080/15391523.2019.1637800.

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.

NWEA. (2019). *MAP® reading fluency™ technical report*.

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.

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.

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*, *48*(9), 923–931. https://doi.org/10.1111/j.1469-7610.2007.01763.x.

Puolakanaho, A., & Latvala, J.-M. (2017). Embedding preschool assessment methods into digital learning games to predict early reading skills. *Human Technology*, *13*(2), 216–236. https://doi.org/10.17011/ht/urn.201711104212.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Rauschenberger, M., Baeza-Yates, R., & Rello, L. (2020). Screening risk of dyslexia through a web-game using language-independent content and machine learning. In Proceedings of the 17th International Web for All Conference, W4A 2020, Taipei Taiwan, 20–21 April 2020. https://doi.org/10.1145/3371300.3383342.

Rauschenberger, M., Baeza–Yates, R., & Rello, L. (2019). Technologies for Dyslexia. In Y. Yesilada & S. Harper (Eds.), *Web accessibility. human–computer interaction series*. Springer. https://doi.org/10.1007/978-1-4471-7440-0_31.

Richardson, U., & Lyytinen, H. (2014). The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology*, *10*(1), 39–60. https://doi.org/10.17011/ht/urn.201405281859.

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.

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*. Basic Books.

Singleton, C., Thomas, K., & Horne, J. (2000). Computer-based cognitive assessment and the development of reading. *Journal of Research in Reading*, *23*(2), 158–180. https://doi.org/10.1111/1467-9817.00112.

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.

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*, *24*(2), 1–18. https://doi.org/10.1080/10888438.2019.1615491.

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.

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.

Wanzek, J., & Vaughn, S. (2007). Research'Based implications from extensive early reading interventions. *School Psychology Review*, *36*(4), 541–561. https://doi.org/10.1080/02796015.2007.12087917.

Watkins, C. P., & Dehaene, S. (2021). *Can a game application that boosts phonics knowledge in kindergarten advance 1st grade reading?* https://doi.org/10.31234/osf.io/pwumg.

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

Zugarramurdi, C., Fernández, L., Lallier, M., Valle-Lisboa, J. C., & Carreiras, M. (2022). Mind the orthography: Revisiting the contribution of pre-reading phonological awareness to reading acquisition. *Developmental Psychology*.

Author Biographies

Camila Zugarramudi is PhD in linguistics, Basque Center on Cognition, Brain and Language and University of the Basque Country (Spain). Msc. in Neuroscience, Universidad de la República-PEDECIBA (Uruguay), Bsc. in Human Biology, Universidad de la República-PEDECIBA (Uruguay). She is an Assistant Professor in the School of Psychology at the Universidad de la República, Uruguay. Her research focuses on language development and reading acquisition from an interdisciplinary perspective, integrating approaches from Cognitive Neuroscience, Developmental Psychology and Education.

Lucía Fernández has a Master's degree in Neuroscience (Programa de Desarrollo de las Ciencias Básicas, PEDECIBA) and a Bachelor in Psychology (Universidad de la República). Her research interests focus on the early stages of reading acquisition. Actually working at Universidad de la República and lecturing at the Universidad Católica. Also works in the assessment and intervention with children with Learning Disabilities and Autistic Spectrum Disorder.

Marie Lallier received her PhD from the University of Grenoble, France. She held post-doc positions in Spain and The United Kingdom. She received the Ramon y Cajal Fellowship and is now a Staff Scientist and Group Leader of the group Educational Neuroscience and Developmental Language Disorders at the Basque Center on

Cognition, Brain and Language. Her main research interests are Language and Reading Acquisition, Developmental Dyslexia, Bilingualism, Attention, and Visual and Auditory processing. She has published more than 45 papers in high impact journals.

Manuel Carreiras is the scientific director of the Basque Center on Cognition, Brain and Language. He is also IKERBASQUE research professor, Honorary Professor of the University College of London, and Distinguished Researcher of the University of the Basque Country. His research focuses on reading, bilingualism, and second language learning. He has published more than 250 papers in high impact journals in the field. His research has been funded by different research agencies. He has been recipient of the Euskadi Research prize 2015, and of the Spanish National Research prize in Social Sciences Pascual Madoz, 2019.

Juan C. Valle-Lisboa, PhD in Biological sciences, Universidad de la República-PEDECIBA (Uruguay). He is researcher of the National System of Researchers and member of the Program for the Development of Basic Science. He currently directs the Interdisciplinary Center of Cognition for Teaching and Learning (CICEA). He is an Aggregate Professor in the School of Sciences at the Universidad de la República, Uruguay. His current research interests include the Neurobiological basis of Language and Cognition, focusing on Neural Network models, field experiments and Neurocognitive studies.