

## Prediction of Reading and Writing in Elementary Education through Early Childhood Education

Talita de Cassia Batista Pazeto<sup>1</sup>

<sup>1</sup>Universidade Presbiteriana Mackenzie, SP, Brazil.

Cristiano Mauro Assis Gomes<sup>3</sup>

<sup>3</sup>Universidade Federal de Minas Gerais, MG, Brazil.

Natália Martins Dias<sup>2</sup>

<sup>2</sup>Universidade Federal de Santa Catarina, SC, Brazil.

Alessandra Gotuzo Seabra<sup>1</sup>

<sup>1</sup>Universidade Presbiteriana Mackenzie, SP, Brazil.

**Abstract:** The study investigated the predictive role of executive functions, language, initial reading and writing abilities, teacher's perception of the students' difficulties and family characteristics, evaluated in Early Childhood Education (ECE) (*Jardim I* and *Jardim II*), in relation to reading and writing performance in the first year of Elementary Education. A total of 71 children were monitored in a longitudinal manner from ECE up to the first year, and evaluated in Oral Language (OL), Executive Functions (EF), initial Reading and Writing skills in ECE and, in the first year, evaluated by means of Reading and Writing tests. Parents provided information on socioeconomic status (SES) and teachers indicated children with difficulties. After the data imputation process, regression tree analysis showed that OL skills and initials skills of reading and writing in ECE explained a mean of 43% of the reading variability in the first year. For writing, the models also included performances in EF, income, father's age and indication of difficulty by the teacher in ECE, explaining a mean of 78% of the variability in writing in the first year. The results allowed identifying abilities and variables that can be considered precursors of subsequent performances in Elementary Education, providing guidance for actions of early identification and intervention.

**Keywords:** Prediction, Written Language, Executive Functions, Evaluation, Oral Language.

---

## Predição de Leitura e Escrita no Ensino Fundamental por meio da Educação infantil

**Resumo:** O estudo investigou o papel preditivo de funções executivas, a linguagem, as habilidades iniciais de leitura e escrita, a percepção do professor e as características familiares sobre as dificuldades de seus alunos, avaliadas na Educação Infantil (EI) (*Jardim I* e *Jardim II*), em relação ao desempenho em leitura e escrita no primeiro ano do Ensino Fundamental. Participaram 71 crianças acompanhadas longitudinalmente no curso da EI em níveis anteriores ao 1º ano, avaliadas em Linguagem Oral (LO), Funções Executivas (FE), habilidades iniciais de Leitura e de Escrita na EI e, no 1º ano, em testes de Leitura e de Escrita. Pais proveram informações sobre nível socioeconômico e professores indicaram crianças com dificuldades. Após processo de imputação de dados, análise de regressão em árvore evidenciou que habilidades de LO e iniciais de leitura e escrita na EI explicaram uma média de 43% da variabilidade em leitura no 1º ano. Para escrita, os modelos incluíram ainda desempenhos em FE, renda, idade do pai e indicação de dificuldade pelo professor na EI, explicando média de 78% da variabilidade em escrita no 1º ano. Os resultados permitiram identificar habilidades e variáveis que podem ser consideradas precursoras para desempenhos posteriores no Ensino Fundamental, podendo informar ações de identificação e intervenção precoces.

**Palavras-chave:** Predição, Linguagem Escrita, Funções Executivas, Avaliação, Linguagem Oral.

## Predicción de la Lectura y Escritura en los Primeros Años de la Educación Infantil

**Resumen:** El estudio investigó el papel predictivo de funciones ejecutivas: lenguaje, habilidades iniciales de lectura y escritura, percepción del profesor y características familiares sobre las dificultades de sus alumnos, evaluadas en la Educación Infantil (*Jardim I y Jardim II*), en relación al desempeño en la lectura y escritura en el primer año de la primaria. Participaron 71 niños acompañados de manera longitudinal desde la Educación Infantil (EI) hasta el 1.º año, quienes fueron evaluados en Lengua Oral (LO), Funciones Ejecutivas (FE), habilidades iniciales de Lectura y de Escritura en la EI y, en el 1.º año los evaluaron en los exámenes de lectura y escritura. Los padres proporcionaron información sobre el nivel socioeconómico, y los profesores indicaron a los niños con dificultades. Después del proceso de imputación de datos, el análisis de regresión en árbol evidenció que las habilidades de LO y de lectura y escritura iniciales en la EI explicaron un promedio del 43% de la variabilidad en lectura en el primer año. Para la escritura, los modelos incluyeron los desempeños en la EI, renta, edad del padre e indicación de dificultad por el profesor en la EI, explicando la media del 78% de la variabilidad en escritura en el primer año. Los resultados permitieron identificar habilidades y variables que pueden ser consideradas precursoras para los desempeños posteriores en la primaria, pudiendo informar acciones de identificación e intervención precoces.

**Palabras clave:** Predicción, Lenguaje Escrito, Funciones Ejecutivas, Evaluación, Lenguaje Oral.

Achieving competence in reading and writing is an important acquisition in the school period and an indispensable tool for learning throughout academic life. This is a complex process that depends on the development of other skills, such as oral language (Melby-Lervåg et al., 2012; Seabra & Dias, 2012a; Song et al., 2015; Torppa et al., 2010), preliminary reading and writing skills (Capovilla & Dias, 2008; Costa et al., 2013), and executive functions (Diamond, 2013; Duncan et al., 2007).

Regarding oral language (OL), this involves different abilities, and, among which, studies have highlighted phonological awareness (PA, ability to reflect on speech sounds), short-term phonological memory (ability to store phonological information for short periods of time) and vocabulary (amplitude of words that the individual knows and/or produces, extension of their semantic lexicon) (e.g., Capovilla & Dias, 2008; Mota et al., 2014; Santos & Ferraz, 2017; Trevisan et al., 2012) as being relevant to the written language.

In spite of evidence about the concurrent relationships between these abilities, few studies, especially in Brazil, have been conducted on their predictive power on reading and writing competences. According to international research, evidence suggests that these

abilities undergo important developments in the years prior to entering formal education at the preschool stage (Suortti & Lipponen, 2016; Vvan Druten-Frietman et al., 2015) and have important predictive power in relation to later reading/writing abilities, especially in the case of PA (Catts et al., 2014; Catts et al., 2015; Melby-Lervåg et al., 2012; Seabra & Dias, 2012a; Song et al., 2015; Torppa et al., 2010).

Preliminary or initial reading and writing (RW) abilities have figured in some studies as being associated with future performance. These abilities include, for example, knowledge of the letters and their sounds, or even the child's ability to encode and/or decode syllables or isolated words (Capovilla & Dias, 2008; Costa et al., 2013; Pazeto et al., 2014; Puranik & Lonigan, 2012). This relationship between the initial competence in written language and the subsequent progress in it can be explained by the Matthew effect, described by Stanovich (1986), in which children with better initial abilities can benefit more from instructions and exposure to the written language and, therefore, develop more, while those with fewer abilities would have fewer opportunities to obtain the same benefit.

In addition, a set of skills known as executive functions (EF) seems to collaborate to written language

acquisition and school performance (Diamond, 2013; Duncan et al., 2007). Executive functions are processes that allow behavioral and cognitive control and include abilities such as inhibition (ability to inhibit inappropriate behaviors or attention to distractors by associating with selective attention), cognitive flexibility (ability to change the attentional or perspective focus), and working memory (ability to sustain and manipulate information mentally) (Diamond, 2013). Although these skills probably have a more important role in more complex processes, such as those involved in reading comprehension (Kendeou et al., 2014; Seabra & Dias, 2012a), they are skills that are also required in learning and the resolution of any new and challenging task (Diamond, 2013), therefore they are expected to have some role in the process of reading and writing acquisition. In fact, even in preschool children, a stage in which these skills are rapidly developing (Center on the Developing Child at Harvard University, 2011), EF are significantly associated with initial RW abilities (Pazeto et al., 2014).

There is also need to consider the dynamic influence of environmental variables on neurodevelopment and, consequently, their impact on learning (Siqueira & Gurgel-Giannetti, 2011), and, among which, variables such as parents' socioeconomic status (SES) and schooling can be highlighted (Enricone & Salles, 2011; Hair et al., 2015; Noble et al., 2012). Thus, individual factors, as cognitive abilities, and extrinsic factors, as environment variables, can contribute to the development of abilities of great relevance in the school, such as reading and writing. Mapping these several variables and assessing cognitive abilities in childhood, especially in preschool children, can help understand how learning occurs and which skills/variables are involved in this process. Another variable that may have some sensitivity in the early detection of children at risk of problems in reading and writing may be the teacher's perception of the child's difficulties in Early Childhood Education. Although studies indicate teacher's good sensitivity (Feitosa et al., 2007; Salles & Parente, 2007), this variable still needs to be tested for its sensitivity at an early age and in a longitudinal cohort.

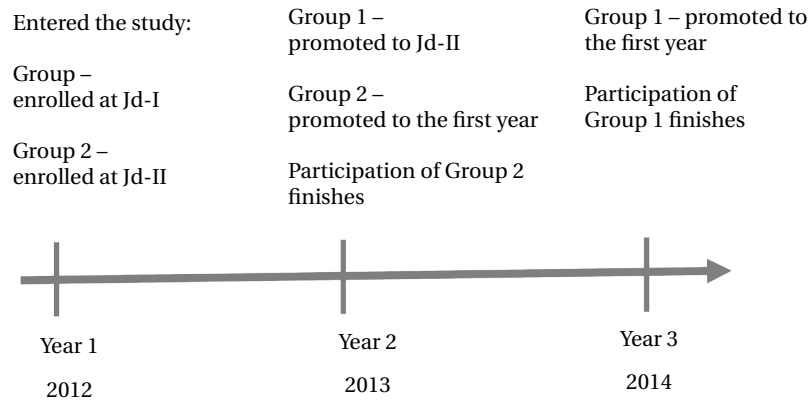
The consideration of all these variables can contribute to the early detection of individuals at risk of school difficulties and help in the planning of appropriate interventions and even public policies aimed at Early Childhood Education. Although there are already studies in the area showing the predictive

relationships between some of the components considered in this study, this research sought to consider all these variables together (cognitive abilities, family and teacher's perception), in a longitudinal cohort, in the prediction of reading and writing in the first year of Elementary Education (EE). Thus, the aim of this study was to investigate predictive models of reading and writing performance in the first year, verifying the influence of abilities evaluated at preschool age (EF, OL, and initial RW abilities) and other family variables (including: income, parental schooling, parents' age, school starting age, childhood illnesses and developmental information, and monitoring by any health or education professional, such as psychologist, speech therapist, etc.). We expected major predictive power of linguistic variables, especially PA and vocabulary, since they represent basic abilities for phonological and semantic aspects of reading and writing; as well as an important predictive power of initial reading and writing measures, because of the Matthew effect, as described earlier. However, EF and familial variables were also expected to have significant prediction power, although in a lower degree.

## Method

### Participants

The sample consisted of 71 children who entered the study during ECE and were followed up to the first year of EE. Thus, for the constitution of this sample, only the children who had participated in at least two moments of the evaluation were retained. These participants entered the study at different periods, so that they formed 2 groups: Group 1, composed of 27 children (age at the beginning of the study: Average ( $A$ ) = 4.4 years; Standard Deviation ( $SD$ ) = 0.50) recruited for the study when they attended *Jardim I* (Jd-I) and who participated in three years of evaluation; and Group 2, composed of 44 children (age at the beginning of the study:  $M$  = 5.3 years;  $SD$  = 0.52) recruited for the study while attending *Jardim II* (Jd-II) and who participated in two years of evaluation. Figure 1 shows the timeline, with the entry of both groups, over the years of the study. The participating school was private and located in an upper middle-class region. The phonic approach was used as the method of literacy and music, computer science and technology education classes were included in their regular grades, which were obligatory for all students.



**Figure 1**  
Time course, with the entry and participation of the two groups, over the three years of the study.

For the constitution of this sample, the exclusion criteria considered were presence of diagnosis of intellectual deficiencies or medical diagnosis of genetic or neurodevelopmental diseases, evidenced by medical report, with this information obtained from the school management and from a questionnaire sent to

the parents/guardians. Four children were excluded for being within these criteria. Any age-grade discrepancy would be considered other exclusion criterion; however, no subject presented this characteristic. Table 1 summarizes the characteristics of the sample for groups 1 and 2.

**Table 1**  
Characterization of the final study sample.

Group	School level at the beginning of the study	Sex	Age at the beginning of the study	Family income	Total number of participants
1	<i>Jardim I</i>	14 M 13 F	4 years old: 16 children 5 years old: 11 children	1 in range 1 3 in range 2 5 in range 3 16 in range 4	27
2	<i>Jardim II</i>	21 M 23 F	5 years old: 31 children 6 years old: 12 children	1 in range 1 8 in range 2 6 in range 3 29 in range 4	44

Note: M = male, F = female/in relation to monthly family income: range 1 = up to R\$ 1,500.00; range 2 = from R\$ 1,500 to R\$ 3,000.00; range 3 = R\$3,000.00 to R\$ 5,000.00; range 4 = Above R\$ 5,000.00 (Note: U.S. dollar/real conversion rate is 3.30, that is, US\$ 1.00=R\$ 3.30).

**Instruments**

*Evaluation of EF (application in Jd-I and Jd-II)*

- **Semantic Stroop Test** (SST; Trevisan, 2010): evaluates the inhibitory control ability and was developed based on the version by Berwid et al. (2005). It is computerized and divided into two parts. In the first part, the child has to name the figures on the screen one by one (boy, girl, moon,

and sun), and in the second he/she has to say the opposite noun of the semantic pair (Example: say “boy” for the figure of the “girl”). Each part has 16 items, presented for a time of 300 ms. Correct responses and reaction time (RT) are recorded in part 2 of the test. The application is individual, with an average duration of 15 minutes. Evidence of validity can be found in Trevisan (2010);

- *Trail Making Test for preschool children* (TT-PS; Trevisan & Seabra, 2012): evaluates the cognitive flexibility ability. In condition A, the child has to connect the stimuli (figures of five puppies) in order of size. In condition B, figures of bones the same sizes as the dogs are introduced, and the child must match the puppies with their appropriate bones, in order of size, connecting them alternately. Performance was measured from the score in sequences (i.e., items connected correctly in an unbroken sequence) in part B of the test (0 to 5 points). The application is individual, with an average duration of 10 minutes. Evidence of validity and normative data have been published for preschool children (Seabra & Dias, 2012a); and
- *Attention by Cancellation Test* (ACT; Seabra & Dias, 2012b): evaluates the selective and alternating attention. In the first part, a target stimulus (a geometric shape) among the other distracters must be cancelled. The same is requested in the second part, however the target stimulus is formed by two geometric shapes. In the last part, the target stimulus changes with each line, requiring the ability to alternate attention. Performance was measured by the total number of correct responses in the test (0 to 132 points). There is a time limit of one minute for each part. The instrument has evidence of validity and standards available (Seabra & Dias, 2012b).

#### *Evaluation of OL (application in Jd-I and Jd-II)*

- *Phonological Awareness Test by Oral Production* (PATO; Seabra & Capovilla, 2012): evaluates the ability to manipulate speech segments. It consists of 10 subtests that measure different levels of phonological awareness: judgment of rhyme and alliteration; synthesis, segmentation, manipulation, and syllabic and phonemic transposition. The application lasts approximately 20 minutes. Evidence of validity, accuracy data and standards have been published by Seabra and Dias (2012b). The total score in the test was used (0 to 40 points);
- *Peabody Picture Vocabulary Test* (PPVT; Capovilla & Capovilla, 1997): evaluates the receptive vocabulary. It is computerized and composed of 125 test items. Each item is composed of four black line drawings on a white background. The program emits a word and the child has to point to the picture that corresponds to

the word being emitted. The application lasts approximately 30 minutes. Evidence of validity and reliability data in a Brazilian children sample were presented by Ferracini et al. (2006). The total score in the test was used; and

- *Word and Pseudoword Repetition Test* (WPRT; Seabra, 2013): evaluates short-term phonological memory. The person applying the test pronounces sequences of two to six words for the child, with the child's task being to repeat the words in the same sequence. Subsequently, sequences with pseudowords are presented. One point is computed for each sequence that is repeated correctly. The total score in the test was used (0 to 20 points). The application lasts approximately 10 minutes. Evidence of validity and normative data were presented by Seabra and Dias (2012b).

#### *Evaluation of initial RW abilities (application in Jd-I and Jd-II)*

- *Letter and sound recognition task* (LSRT; Pazeto et al., 2014): evaluates the child's knowledge of the names of the letters and their sounds. Uppercase letters are randomly presented to the child on printed sheets, one at a time. In the first application, the child has to name the letter; then the procedure is repeated, however, the child is instructed to say the sound of the letter. Total scores in each part were used, which could range from 0 to 26 points. The application lasts approximately 10 minutes. A national study has already used the task in a preschool children sample (Pazeto et al., 2014); and
- *Reading and Writing Task* (RWT; Pazeto et al., 2014): assesses preliminary reading and writing abilities. It consists of two phases: in the first, the child receives a sheet with eight words and two pseudowords to read, and in the second, eight words and two pseudowords are dictated for the child to write. The items are classified as regular high and low frequency, irregular high and low frequency, and pseudowords, with two items of each type. The percentage of correct responses per item is calculated, varying from 0 to 100% in each part of the test, depending on the number of correct letters in relation to the total letters of the item. The application lasts approximately 10 minutes. A national study has already used RWT in a preschool children sample (Pazeto et al., 2014).

*Assessment of reading and writing (application in the first year of EE)*

- *Word-Pseudoword Reading Competence Test* (WPRCT; Seabra & Capovilla, 2010): evaluates the recognition of isolated words. It has 70 items, each consisting of a figure and a written element, which can be correctly related to the figure (two types: graphophonetically regular or irregular correct word), or may be incorrect (five types: semantic, visual, phonological, spelling error, or strange error in which there is no relation to the figure). The objective is to judge whether the written word corresponds to the figure. The application lasts approximately 30 minutes. Data on the psychometric and normative characteristics are provided in the test manual. The total score in the test was used (0 to 70 points);
- *Writing under Dictation Test – reduced version* (WDT-rv; Seabra & Capovilla, 2013): evaluates writing in the dictation condition. The applicator pronounces 36 psycholinguistic items, one by one, and the child has to write them on a lined sheet of paper. All the items belong to the list provided by Pinheiro (1994) and vary in terms of lexicality, regularity of graphophonic correspondences, their frequency of occurrence in the Brazilian Portuguese language, and their length. The application lasts approximately 30 minutes. The total number of errors in the test was used (to compute the errors, the sum of all graphemes recorded incorrectly in each of the 36 items of the PED-vr is considered, which can vary from 0 to 210). Psychometric characteristics and normative data are presented by Seabra, Dias and Capovilla (2013).

*Questionnaires (application in the year the child entered the study, Jd-I or Jd-II)*

- *Questionnaire for Parent Identification (Qpa)*: asks about important aspects of the child, his/her history, development and health, as well as information for characterization of the family, to be completed by the parents. The completion time is approximately five minutes. The variables selected for analysis were: prematurity; school starting age; mother's age and schooling; father's age and schooling; family income; and
- *Questionnaire for Teachers (Qtch)*: The teachers were asked to indicate which students, in their judgment, had some type of difficulty (without

specifying whether it was academic or behavioral) when compared to their class peers. The responses were annotated to verify the participation of this variable in the prediction models.

## Procedure

The project was approved by the Research Ethics Committee (CAAE 02631312.3.0000.0084), and after its approval contact was made with the school and the consent forms sent by the institution to those responsible for the children. After the signing of the consent form, data collection was started, which occurred in the school over three years, always in the second semester, between mid-August and mid-October. The children, during the course of Jd-I and Jd-II, were taken from the room individually, in sessions of 10 to 30 minutes, with the consent from the teacher and student. Each participant was taken out at least six times to apply the tests, individually. When in the first year, WPRCT and WDT tests were applied in a collective manner, in the class, with an average duration of 40 minutes, in two sessions. In addition, parents responded to the Qpa and teachers to the Qtch at the time the child entered the study (Jd-I and Jd-II for groups 1 and 2, respectively).

## Data analysis

*Normality check and database preparation step:* The data underwent analysis of the parameters of normality and the distributions were not normal in most of measurements, indicating non-parametric statistical analyzes. In order to prepare the database for the analysis of reading and writing performance predictions, the data were imputed. Imputation is a statistical procedure used in studies with missing data, aiming to complete the data. This procedure can be used in longitudinal studies, allowing predictive analyzes to be conducted more robustly (Carpenter & Kenward, 2013). It was performed using the R Software, version 3.2.2 (R Core Team, 2015). The procedure starts from the number of initial subjects and derives a value for the missing results of each subject, so that, at the end of the imputation, the database is complete. This process was repeated five times, creating five complete databases. According to Carpenter and Kenward (2013), this number of imputations is sufficient to obtain acceptable properties and to guarantee greater effectiveness and reliability

of the data. In the imputation, the data created are obtained through an analysis of the child's own results in the other years and the other children in the same year and test, extracting an expected pattern for that subject in that specific ability (Carpenter & Kenward, 2013). The process enabled the preparation of the database for the subsequent analyses. All the imputations generated presented Rhat values lower than 1.1, indicating that the data estimated for the five imputed worksheets can be used for the other analyses.

*Analyses focused on the study objective:* Linear Regression Tree Analysis, performed using the R Software through M5 algorithm, was conducted to verify which cognitive abilities and other variables, measured in ECE, acted as predictors of reading and writing performance in the first year of EE. The following predictor variables were used: for executive functions, Stroop Test score as measurement of inhibitory control, Attention by Cancellation Test score as measurement of attention, and Trail Making Test for preschool children score as measurement of cognitive flexibility; for language, PATO score as measurement of phonological awareness, Peabody Picture Vocabulary Test (PPVT) score as measurement of receptive vocabulary, and Word and Pseudowords Repetition Test score as measurement of phonological memory; for preliminary reading and writing skills, Knowledge of Letters and Sounds score as measurement of recognition of letters and sounds, and Reading and Writing Test score as measurement of word recognition and ability to read and write; for family variables, income, parental schooling, parents' age, school starting age, and childhood illnesses; and for developmental information, monitoring by any health or education professional, such as psychologist, speech therapist, among others.

The regression tree analysis produces models of the importance of each variable for the prediction of the outcome variable. The "leave-one-out

cross-validation" method was used, which is a control strategy that aims to minimize overfitting. This method has the advantage of being non-parametric and non-linear. According to Golino and Gomes (2014), such methods are especially interesting for the area of educational research because they allow analyzing which set of variables best predicts a certain outcome. In this study, such an analysis was conducted from the general worksheet derived from the five imputations performed.

In order to select the independent variables that actually had incremental validity for the explanation of the dependent variables, the Boruta package (Kursa & Rudnicki, 2010) was used, which is a R Software package that performs, through the Random Forest approach, the selection of variables to compose a predictive model to be tested (Kursa & Rudnicki, 2010). This process allowed more parsimonious models to be produced, with predictor variables selected according to their importance in the model. For each model, RMSE (*root mean square error*) was used, which indicates how much the sample is not able to explain the outcome variable, used only to compare the samples (does not have a maximum or minimum value), and  $R^2$ , used as an index of fit of the model, referring to the variability of the outcome variable explained by the variability of the predictor variables.

## Results

### Prediction of the Reading Performance

The analysis was conducted using the Boruta package (Kursa & Rudnicki, 2010), in order to verify which independent variables would have incremental validity to explain the outcome variable. This analysis was performed for each of the five samples generated by the imputation process. These variables are presented in Table 2 in order of importance in predicting the outcome.

Table 2

Variables confirmed by Boruta in the prediction of Reading and results of the regression analysis by imputation sample and confidence intervals for the prediction of Reading.

Samples	Ai (1)	Ai (2)	Ai (3)	Ai (4)	Ai (5)		
Variables Confirmed (Boruta)	KnoLJd2	PatoJd2	PatoJd2	PatoJd2	PatoJd2		
	PatoJd2	PpvtJd2	PpvtJd2	KnoLJd2	KnoLJd2		
	PpvtJd2	KnoLJd2	KnoLJd2	PpvtJd2	PpvtJd2		
	ReadJd2	ReadJd2	WritJd2	WritJd2	WritJd2		
	WritJd2	WritJd2	ReadJd2	ReadJd2	ReadJd2		
	PatoJd1	KnoSjd2	KnoSjd2	KnoSjd2	ActTCRJd1		
	KnoSjd2	*	*	*	KnoSjd2		
Results of Regression Tree Analysis							
	Ai (1)	Ai (2)	Ai (3)	Ai (4)	Ai (5)	R <sup>2</sup>	
						M	SD
RMSE	4.98	5.64	5.64	5.64	5.64	--	--
R <sup>2</sup>	0.53	0.40	0.40	0.40	0.40	0.43	0.06
CI (Min-Max)	0.30-0.64	0.16-0.53	0.16-0.53	0.16-0.53	0.15-0.53	--	--

KnoLJd2: Knowledge of Letters Task (*Jardim* II); KnoSjd2: Knowledge of Sounds Task (*Jardim* II); WritJd2: Writing Test (*Jardim* II); ReadJd2: Reading Test (*Jardim* II); PatoJd1: Phonological Awareness Test by Oral Production (*Jardim* I); PatoJd2: Phonological Awareness Test by Oral Production (*Jardim* II); ActTCRJd1: Attention Cancellation Test total correct responses (*Jardim* I).

Between six and seven variables were maintained in the models, with the performances in Jd-II in phonological awareness; in the knowledge of letters and sounds task; performance in RWT, both in writing and in reading; and performance in PPVT (which appeared in the models in all five imputations) being the most important; and, in Jd-I, the performances in phonological awareness (appeared in one imputation) and in attention (one imputation). Table 2 shows the regression analysis values, including RMSE and R<sup>2</sup>. It also presents the 95% confidence interval (CI) values for each of the samples imputed in the prediction of reading. Considering all five samples generated by the imputation, the independent variables selected by the Boruta package explained a mean of 43% of the variance of reading performance in the first year of elementary education.

### Prediction of the Writing Performance

Boruta analysis was conducted to confirm which variables were significant for each imputed sample.

These variables are presented in Table 3 in order of importance in predicting the outcome.

In all the generated models, 10 variables were maintained, the most important of which were performance in Knowledge of letters in Jd-II; difficulty highlighted by the teacher; performances in the RWT in Jd-II, both in writing and in reading; performances in the phonological awareness and vocabulary tests in Jd-II; father's age; income; and performance in the ACT in the Jd-II (all in the five imputations). In addition, the variables performance in the SST in Jd-II (three imputations); participation group (one imputation) and performance in the WPRT in Jd-I (one imputation) were also included in some models. Table 3 presents the regression analysis values, including RMSE and R<sup>2</sup>, also for the prediction of writing. It also presents CI values for each of the imputed samples. Considering all five samples generated by the imputation, the independent variables selected explained a mean of 78% of the variance of writing performance in the first year of elementary education.



Table 3

Variables confirmed by Boruta in the prediction of Writing and results of the regression analysis by imputation sample and confidence intervals for the prediction of Writing

Samples	Ai (1)	Ai (2)	Ai (3)	Ai (4)	Ai (5)		
Variables Confirmed (Boruta)	WritJd2	Difficulty	Income	Difficulty	Difficulty		
	Difficulty	WritJd2	Difficulty	WritJd2	WritJd2		
	Income	ActTCRJd2	WritJd2	Income	ActTCRJd2		
	ActTCRJd2	Income	PpvtJd2	ActTCRJd2	Income		
	KnoLJd2	PpvtJd2	PatoJd2	ReadJd2	PpvtJd2		
	PatoJd2	ReadJd2	ReadJd2	PatoJd2	ReadJd2		
	PpvtJd2	KnoLJd2	ActTCRJd2	PpvtJd2	PatoJd2		
	ReadJd2	Father's age	Father's age	Father's age	Father's age		
	Father's age	PatoJd2	KnoLJd2	KnoLJd2	KnoLJd2		
	RttJd1	Group	St2CRJd2	St2CRJd2	St2CRJd2		
Results of the Regression Tree Analysis							
	Ai (1)	Ai (2)	Ai (3)	Ai (4)	Ai (5)	R <sup>2</sup>	
						M	SD
RMSE	11.55	11.25	11.35	11.33	11.05	--	--
R <sup>2</sup>	0.77	0.79	0.78	0.78	0.79	0.78	0.008
CI (Min-Max)	0.60-0.83	0.62-0.84	0.62-0.84	0.62-0.84	0.63-0.84	--	--

KnoLJd2: Knowledge of Letters Task (*Jardim II*); Difficulty: Child with difficulty highlighted by the teacher; WritJd2: Writing Test (*Jardim II*); Group: Group according to the number of years the child participated in the study; Father's age: Father's age when the child entered the study; ReadJd2: Reading Test (*Jardim II*); PatoJd2: Phonological Awareness Test by Oral Production (*Jardim II*); Income: Family income range; St2CRJd2: Semantic Stroop test part 2 correct responses (*Jardim II*); ActTCRJd2: Attention Cancellation Test, total correct responses (*Jardim II*); RttJd1: Repeat Test Total (*Jardim I*); PpvtJd2: Peabody Picture Vocabulary Test (*Jardim II*).

## Discussion

The aim of the study was to test predictive models of reading and writing performance in the first year of Elementary Education, verifying the influence of the abilities assessed at preschool age (EF, OL and initial RW abilities) and other variables (family and teacher's perspective). With regard to the reading prediction, it was found that phonological awareness, knowledge of letters and sounds, initial RW and vocabulary abilities, all in Jd-II, were the main predictive abilities for later reading performance. Only one model in an imputed sample again included performance in phonological awareness, however in Jd-I, and performance in attention, also in Jd-I. Thus, there was no consistency in the selection of these two variables, which seem not to be as relevant as the first ones. The models generated explained a mean of 43% of the variance in the reading performance in the first year.

Considering that written language is a secondary representation of oral language (Trevisan et al., 2012),

the participation of linguistic skills in the models was expected, in addition to knowledge of letters and sounds, which are preliminary RW skills. The relationship between phonological awareness and reading is already well established in the literature, with evidence that this ability is the best predictor of future reading performance in the Portuguese Language (Capovilla & Dias, 2008; Carvalhais, 2011; Seabra & Dias, 2012c; Trevisan et al., 2012) and in other alphabetic orthographies (Catts et al., 2014; Catts et al., 2015; Foorman et al., 2015; Melby-Lervåg et al., 2012).

The knowledge of letters and sounds has also been highlighted as relevant to reading (Capovilla & Dias, 2008; Costa et al., 2013), and participation of initial RW abilities in predictive reading models can be explained by the Matthew effect phenomenon (Stanovich, 1986). There is also evidence supporting the participation of vocabulary in the explanatory models. Torppa et al. (2010) found that children with difficulty in the reading area already differed in

language skills, including vocabulary, at two years of age, reinforcing the importance of this ability at early ages. More recently, Song et al. (2015) found that vocabulary development, assessed over an 8-year study, could explain subsequent reading skills.

With regard to the writing prediction, it was verified that knowledge of letters, initial reading and writing abilities, AT, vocabulary and attention, in Jd II, were the main predictor skills for subsequent writing performance, as well as variables such as difficulty highlighted by the teacher, father's age, and income. Three of the imputed samples included the performance in the test that evaluates inhibition. The participation group variable and performance in WPRT (phonological memory) were only selected in one imputed sample, in Jd-I, so that there was no consistency in the selection of these variables, which seem not to be as relevant as the first ones. The models generated explained a mean of 78% of the variance in the writing performance in the first year.

There is a reasonable overlap between some skills that were included in the reading and writing models, such as phonological awareness, vocabulary, knowledge of letters, and initial reading and writing skills. Although these variables are more approached in the literature in relation to reading, the same understanding regarding their participation in explanatory models can be applied to writing, with some studies having also been devoted to this specific association (Carvalhais, 2011; França et al., 2004), even in preschool children (Pazeto et al., 2014; Pestun, 2005).

In addition to the linguistic and emerging reading and writing abilities, others such as attention and inhibition were selected in the prediction models, suggesting their contribution in the development of writing. Pazeto et al. (2014) also found correlations between EF abilities, such as attention, cognitive flexibility and inhibitory control, and writing in preschool children aged four and five years. Duncan et al. (2007), when analyzing six longitudinal studies, also observed that EF predict academic success. For Diamond (2013), inhibitory control is related, for example, to the discipline to remain on a task and involves controlling the focus of attention to distractors, skills also required in writing and, according to the author, part of the daily school routine.

Other variables were included in the explanatory models generated, in addition to the cognitive abilities evaluated, such as difficulty highlighted by the teacher,

father's age, and income. These results suggest that the teacher's judgment as to the presence or absence of difficulty (of learning or behavior) in the child seems to be sensitive in predicting future performance in writing. In fact, Salles and Parente (2007) stated that the teacher's evaluation (in their study, of the children's reading and writing ability) was presented as a sensitive factor in relation to the students' performance.

The father's age also figured as a predictor of writing, in that the greater the father's age, the better the child's score. Enricone and Salles (2011) emphasize that family aspects influence the development of school skills such as writing, but do not present the father's age as one of these predictors. One hypothesis is that the father's age may be related to the improvement of other family conditions, such as schooling, income, or even stability in terms of housing, with income also figuring as an important variable for subsequent writing performance (Enricone & Salles, 2011). In general, SES has been shown to be associated with cognitive development and even structural differences in the brain (Hair et al., 2015; Noble et al., 2012), so that its impact on a school performance ability, as in the case of writing, seems to be supported by the literature.

It is worth mentioning that the models generated had higher prediction power for writing ( $R^2$  mean of 78%) than for reading ( $R^2$  mean of 43%). This may have occurred because of greater variability of performance of the sample in writing in relation to reading. It may also be linked to the complexity of the tasks proposed for the evaluation (the reading task, for example, allowed for correct responses at random) or even the abilities involved. For example, in the report *Alfabetização infantil: os novos caminhos* (Children's Literacy: the new paths) (Brazil, 2007), the authors affirm that writing involves the development of additional skills, such as accurate and complete knowledge of orthographic representations, therefore, being a more complex acquisition than reading. If this hypothesis is correct, it would explain the greater demands involved in writing in relation to reading.

The study has some limitations, so some care needs to be taken in interpreting the results. Despite the statistical techniques used to minimize the sample difficulties, the small number of participants that could be monitored during the three years of the study should be noted, mainly due to school changes in the transition from ECE to EE. Other limitations refer to the absence

of an assessment of intelligence, a variable that could influence the predictions made, as well as the fact that the research was carried out exclusively with a sample from a private school, so that the findings may not be representative of the reality of Brazilian public schools. It is hoped that further studies can address these limitations and amplify or even confront these findings.

Despite the limitations, it is considered that the study advances the knowledge available in the area, which is mostly the result of cross-sectional studies, especially due to the lack of Brazilian research. Thus, in a three-year longitudinal cohort, the study illustrated the role of EF, OL, and initial RW in ECE for reading prediction, as well as their role, in conjunction with SES variables and the teacher's perception in the prediction of writing in the first year of EE. In terms of implications, the study highlights the role of assessment of these skills and consideration of these variables in the early identification of children at risk of poor reading and/or writing performance in order to provide appropriate interventions and monitor

their progress in the course of ECE, aiming to minimize future difficulties when entering EE.

### Final considerations

The results allowed the identification of some abilities and variables that can be considered as precursors of later performances in EE in areas that are greatly needed and considered as bases of development and academic life, that is, reading and writing. In this sense, the results can support reflections on possibilities of action for the early identification of possible problems in these abilities. This, in turn, should encourage the implementation of preventive interventions. Therefore, the study has relevance also for educational policies, since it contributes by highlighting relevant areas to be considered and even developed (supported by the literature and experimental studies) in ECE. This knowledge can be used in the structuring of the curriculum and teacher training, for example, focusing on content that can be worked on and stimulated in children at an early age.

### References

- Berwid, O. G., Curko Kera, E. A., Marks, D. J., Santra, A., Bender, H. A., & Halperin, J. M. (2005). Sustained attention and response inhibition in young children at risk for Attention Deficit/Hyperactivity Disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 46(11), 1219-1229. <https://doi.org/10.1111/j.1469-7610.2005.00417.x>
- Brasil. (2007). Grupo de trabalho alfabetização infantil: os novos caminhos: relatório final. Câmara dos Deputados.
- Capovilla, F. C., & Capovilla, A. G. S. (1997). Desenvolvimento linguístico na criança dos dois aos seis anos: tradução e standardização do PPVT de Dunn & Dunn, e da Language Development Survey de Rescorla. *Ciência Cognitiva*, 1(1), 353-380.
- Capovilla, A. G. S., & Dias, N. M. (2008). Habilidades de linguagem oral e sua contribuição para a posterior aquisição de leitura, *Psic*, 9(2), 135-144.
- Carpenter, J. R., & Kenward, M. G. (2013). *Multiple imputation and its application*. John Bile & Sons.
- Carvalhais, L. (2011). Treino de consciência fonológica em crianças com dificuldade de aprendizagem. *Estratégias para uma Aprendizagem com Sucesso*, 2(1), 1-9.
- Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing*, 28(9), 1407-1425.
- Catts, H. W., Fey, M. E., Weismer, S. E., Bridges, M. S., Tomblin, J. B., & Nippold, M. A. (2014). The relationship between language and reading abilities. In Tomblin B. & Nippold, N. A (Eds.), *Understanding individual differences in language development across the school years* (pp. 144-165). Psychology Press.
- Center on the Developing Child at Harvard University. (2011). *Building the brain's "air traffic control" system: how early experiences shape the development of executive function* [Working paper]. <https://developingchild.harvard.edu/wp-content/uploads/2011/05/How-Early-Experiences-Shape-the-Development-of-Executive-Function.pdf>
- Costa, H. C., Perdry, H., Soria, C., Pulgar, S., Cusin, F., & Dellatolas, G. (2013). Emergent literacy skills, behavior problems and familial antecedents of reading difficulties: a follow-up study of reading achievement from kindergarten to fifth grade. *Research in Developmental Disabilities*, 34(3), 1018-1035. <https://doi.org/10.1016/j.ridd.2012.11.029>

- Diamond, A. (2013). Executive functions. *Annual Reviews*, 64, 135-168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, L., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428-1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- Enricone, J. R. B., & Salles, J. F. (2011). Relação entre variáveis psicossociais familiares e desempenho em leitura/escrita em crianças. *Psicologia Escolar e Educacional*, 15(2), 199-210. <https://doi.org/10.1590/S1413-85572011000200002>
- Feitosa, F. B., Prette, Z. A., & Loureiro, S. R. (2007). Acuracidade do professor na identificação de alunos com dificuldade de aprendizagem. *Temas Em Psicologia*, 15(2), 237-247.
- Ferracini, F., Capovilla, A., Dias, N., & Capovilla, F. (2006). Avaliação de vocabulário expressivo e receptivo na educação infantil. *Psicopedagogia*, 23(71), 124-133.
- Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in Kindergarten through Grade 2. *Reading and writing*, 28(5), 655-681.
- França, M. P., Wolff, C. P.; Moojen, S. & Rotta, N. (2004). Aquisição da Linguagem oral: relação e risco para linguagem escrita. *Arquivos de Neuropsiquiatria*, 62(2b), 469-472.
- Golino, H. F., & Gomes, C. M. A. (2014). Four machine learning methods to predict academic achievement of college students: a comparasion study. *Revista E-Psi*, 4(1), 68-101.
- Hair, N. L., Hanson, J. L., Wolfe, B. L., & Pollak, S. D. (2015). Association of child poverty, brain development, and academic achievement. *JAMA Pediatrics*, 169(9), 822-829. <https://doi.org/10.1001/jamapediatrics.2015.1475>
- Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A Cognitive View of Reading Comprehension: Implications for Reading Difficulties. *Learning Disabilities Research & Practice*, 29(1), 10-16. <https://doi.org/10.1111/ldrp.12025>
- Kursa, B. M., & Rudnicki, W. R. (2010). Feature selection with the boruta package. *Journal of Statistical Software*, 36(11), 1-13.
- 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>
- Mota, M. M. P. E., Santos, A. A. A., Guimarães, S. B., & Conti, C. (2014). Evidências de validade do roteiro de avaliação da consciência fonológica (RACF). *Estudos e Pesquisas em Psicologia*, 14(3), 933-948.
- Noble, K. G., Houston, S. M., Kan, E., & Sowell, E. R. (2012). Neural correlates of socioeconomic status in the developing human brain. *Developmental Science*, 15(4), 516-527. <https://doi.org/10.1111/j.1467-7687.2012.01147.x>
- Pazeto, T. C. B., Seabra, A. G., & Dias, N. M. (2014). Executive functions, oral language and writing in preschool children: Development and correlations. *Paidéia*, 24(58), 213-221. <https://doi.org/10.1590/1982-43272458201409>
- Pestun, M. S. V. (2005). Consciência fonológica no início da escolarização e o desempenho ulterior em leitura e escrita: estudo correlacional. *Estudos em psicologia*, 10(3), 407-412.
- Pinheiro, A. M. V. (1994). *Leitura e escrita: uma abordagem cognitiva*. Editora Psy II.
- Puranik, C. S., & Lonigan, C. J. (2012). Name-writing proficiency, not length of name, is associated with preschool children's emergent literacy skills. *Early Childhood Research Quarterly*, 27(2), 284-294. <https://doi.org/10.1016/j.ecresq.2011.09.003>
- R Core Team. R. (2015). The R Project for statistical computing [Homepage]. <https://www.R-project.org>
- Salles, J. F., & Parente, M. A. M. P. (2007). Relação entre desempenho infantil em linguagem escrita e percepção do professor. *Cadernos de pesquisa*, 37(132), 687-709.
- Santos, A. A. A., & Ferraz, A. S. (2017). Avaliação de habilidades relacionadas à compreensão de leitura no ensino fundamental. *Psico*, 48(1), 21-30. <https://dx.doi.org/10.15448/1980-8623.2017.1.24376>
- Seabra, A. G. (2013). Teste de repetição de palavras e pseudopalavras. In A. Seabra & N. Dias (Eds.), *Avaliação neuropsicológica cognitiva: linguagem oral* (pp. 97-99). Memnon.
- Seabra, A. G., & Capovilla, F. C. (2010). *Teste de competência de leitura de palavras e pseudopalavras (TCLPP)*. Memnon.
- Seabra, A. G., & Capovilla, F. C. (2012). Prova de Consciência Fonológica por produção oral. In A. G. Seabra & N. M. Dias (Eds.), *Avaliação neuropsicológica cognitiva: linguagem oral* (pp. 117-122). Memnon.


- Seabra, A. G., & Capovilla, F. C. (2013). Prova de Escrita sob Ditado: versão reduzida. In A. G. Seabra, N. M. Dias, & F. C. Capovilla (Eds.), *Avaliação Neuropsicológica Cognitiva: leitura, escrita e aritmética* (pp. 70-73). São Paulo: Memnon.
- Seabra, A. G., & Dias, N. M. (2012a). Reconhecimento de palavras e compreensão de leitura: dissociação e habilidades linguístico-mnemônicas preditoras. *Revista Neuropsicologia Latinoamericana*, 4(1), 43-56. <https://dx.doi.org/10.5579/rnl.2012.0101>
- Seabra, A. G., & Dias, N. M. (2012b). *Avaliação neuropsicológica cognitiva: atenção e funções executivas*. Memnon.
- Seabra, A. G., & Dias, N. M. (2012c). *Avaliação neuropsicológica cognitiva: linguagem oral*. Memnon.
- Seabra, A. G., Dias, N. M., & Capovilla, F. C. (2013). *Avaliação neuropsicológica cognitiva: leitura, escrita e aritmética*. Memnon.
- Siqueira, C. M., & Gurgel-Giannetti, J. (2011). Poor school performance: an updated review. *Revista da Associação Médica Brasileira*, 57(1), 78-87. [https://dx.doi.org/10.1016/S2255-4823\(11\)70021-2](https://dx.doi.org/10.1016/S2255-4823(11)70021-2)
- Song, S., Su, M., Kang, C., Liu, H., Zhang, Y., McBride-Chang, C., Tardif, T., Li, H., Liang, W., Zhang, Z., & Shu, H. (2015). Tracing children's vocabulary development from preschool through the school-age years: an 8-year longitudinal study. *Developmental Science*, 18(1), 119-131. <https://dx.doi.org/10.1111/desc.12190>
- Stanovich, K. E. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-406.
- Suortti, O., & Lipponen, L. (2016). Phonological awareness and emerging reading skills of two- to five-year-old children. *Early Child Development and Care*, 186(11), 1703-1721. <https://dx.doi.org/10.1080/03004430.2015.1126832>
- Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., & Lyytinen, H. (2010). Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *Journal of Learning Disabilities*, 43(4), 308-321. <https://dx.doi.org/10.1177/0022219410369096>
- Trevisan, B. T. (2010). *Atenção e controle inibitório em pré-escolares e correlação com indicadores de desatenção e hiperatividade* (Dissertação de mestrado, Universidade Presbiteriana Mackenzie).
- Trevisan, B. T., & Seabra, A. G. (2012). Teste de trilhas para pré-escolares. In A. G. Seabra & N. M. Dias (Eds.), *Avaliação neuropsicológica cognitiva: atenção e funções executivas* (pp. 92-100). Memnon.
- Trevisan, B., Hipólito, R., Martoni, A., Ferracini, F., Dias, N., & Seabra, A. (2012). Teoria e pesquisa para avaliação de aspectos da linguagem oral. In Seabra, A. & Dias, N. (Orgs), *Avaliação neuropsicológica cognitiva: linguagem oral* (pp. 14-23). Memnon.
- Vvan Druten-Frietman, L., Denessen, E., Gijssel, M., & Verhoeven, L. (2015). Child, home and institutional predictors of preschool vocabulary growth. *Learning and Individual Differences*, 43, 92-99. <https://dx.doi.org/10.1016/j.lindif.2015.08.032>

---

*Talita de Cassia Batista Pazeto*

Universidade Presbiteriana Mackenzie, São Paulo – SP, Brasil.


E-mail: talita.psicopedagoga@hotmail.com

 <https://orcid.org/0000-0003-3663-2205>

*Natália Martins Dias*

Departamento de Psicologia da Universidade Federal de Santa Catarina, Santa Catarina – SC, Brasil.


Email: natalia.m.dias@ufsc.br

 <https://orcid.org/0000-0003-1144-5657>

*Cristiano Mauro Assis Gomes*

Professor Associado da Universidade Federal de Minas Gerais, Minas Gerais – MG, Brasil.


Email: cristianomaurogomes@gmail.com

 <https://orcid.org/0000-0003-3939-5807>

*Alessandra Gotuzo Seabra*

Programa de Pós-graduação em Distúrbios do Desenvolvimento pela Universidade Presbiteriana Mackenzie, São Paulo – SP, Brasil.

Email: [alessandragseabra@gmail.com](mailto:alessandragseabra@gmail.com)

 <https://orcid.org/0000-0002-8373-7897>

Endereço para envio de correspondência:

Av. Interlagos, 800, Jardim Marajoara. CEP: 04660-000. São Paulo – SP. Brasil.

*Received* 26/06/2018

*Approved* 25/04/2019

*Recebido* 26/06/2018

*Aceito* 25/04/2019

*Recibido* 26/06/2018

*Aceptado* 25/04/2019

Acknowledgments: CNPq (Productivity scholarship for NM Dias, CMA Gomes & AG Seabra).

*How to cite:* Pazeto, T. C. B., Dias, N. M., Gomes, C. M. A. & Seabra A. G. (2020). Prediction of Reading and Writing in Elementary Education through Early Childhood Education: Reading and Writing Prediction. *Psicologia: Ciência e Profissão*, 40, 1-14. <https://doi.org/10.1590/1982-3703003205497>

*Como citar:* Pazeto, T. C. B., Dias, N. M., Gomes, C. M. A. & Seabra A. G. (2020). Predição de leitura e escrita no ensino fundamental por meio da educação infantil: predição de leitura e escrita. *Psicologia: Ciência e Profissão*, 40, 1-14. <https://doi.org/10.1590/1982-3703003205497>

*Cómo citar:* Pazeto, T. C. B., Dias, N. M., Gomes, C. M. A. & Seabra A. G. (2020). Predicción de la Lectura y Escritura en los Primeros Años de la Educación Infantil: Predicción de la Lectura y Escritura. *Psicologia: Ciência e Profissão*, 40, 1-14. <https://doi.org/10.1590/1982-3703003205497>