Cognition 182 (2019) 1–7

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Statistical learning and spelling: Evidence from Brazilian prephonological spellers☆

Rebecca Treiman[a,*], Cláudia Cardoso-Martins[b], Tatiana Cury Pollo[c], Brett Kessler[a]

[a] Department of Psychological and Brain Sciences, Washington University in St. Louis, United States
[b] Departamento de Psicologia, Universidade Federal de Minas Gerais, Brazil
[c] Departamento de Psicologia, Universidade Federal de São João Del–Rei, Brazil

ARTICLE INFO

Keywords:
Statistical learning
Graphotactics
Spelling
Writing
Literacy development
Portuguese

ABSTRACT

We analyzed the spelling attempts of Brazilian children (age 3 years, 3 months to 6 years, 0 months) who were prephonological spellers, in that they wrote using letters that did not reflect the phonemes in the words they were asked to spell. We tested the hypothesis that children use their statistical-learning skills to learn about the appearance of writing and that older prephonological spellers, who have had on average more exposure to writing, produce more wordlike spellings than younger prephonological spellers. We found that older prephonological spellers produced longer spellings and were more likely to use letters and digrams in proportion to their frequency of occurrence in Portuguese. There were also some age-related differences in children's tendency to use letters from their own names when writing other words. The results extend previous findings with learners of English to children who are learning a more transparent orthography.

1. Introduction

The world is filled with patterns. People take advantage of these patterns by learning how often events occur and how often and under what circumstances they occur together. In this way, they develop a kind of mental statistics that helps them to deal with the environment. For example, someone who hears the word *abscond* for the first time may be able to determine its grammatical category based on the knowledge that two-syllable English words with stress on the second syllable are often verbs. Some patterns are explicitly taught, but others—such as this pattern about word stress—are not. In such cases, people pick up the patterns through implicit statistical learning.

Most studies of statistical learning have examined people's learning of artificial materials over short periods of time. For example, Chambers, Onishi, and Fisher (2003) constructed sets of spoken syllables in which some consonants occurred in the initial position but not the final position and other consonants showed the reverse pattern. Twenty-five such syllables were repeatedly played to infants over the course of 3–4 min. Infants' knowledge of the patterns was then tested by comparing the amount of time they spent listening to new syllables that conformed to the patterns and new syllables that did not conform. The results of this and other studies (e.g., Saffran, Aslin, & Newport, 1996) show that statistical learning is available even to infants and provide some information about its properties. However, such studies may not fully capture statistical learning as it occurs in real life, where patterns may be probabilistic rather than all or none, exposure may be spaced rather than massed, the number of items to which learners are exposed may be large rather than small, and the delay between learning and testing may be long rather than short.

In the present study, we examined a real-life case of statistical learning: learning about the letter patterns in written words. The words of a language have characteristic lengths and letter patterns. In the written words of Portuguese, for example, the letters ‹a› and ‹o› are more common than the letters ‹e› and ‹i›. Among *digrams* (sequences of two adjacent letters), ‹ci› is more common than ‹cr›. Modern children see many examples of writing in their daily lives, giving them an opportunity to learn about such *graphotactic* patterns. In the present study, we assessed Brazilian preschoolers' knowledge about these patterns as reflected in their attempts to write words. Our main analyses involved *prephonological spellers*: children who use letters when asked to write, but not letters that make sense based on the sounds in the target items. Such a child might write ‹cicio› for *tartaruga* 'turtle', for example.

Despite their lack of knowledge of sound–letter relations, prephonological spellers appear to possess some knowledge about the letter patterns of their written language. For example, these children tend to use letters and digrams in proportion to the frequency with

☆ This research was supported by NSF grant BCS-1421279 and NIH grant HD051610.
* Corresponding author at: Department of Psychological and Brain Sciences, Campus Box 1125, St. Louis, MO 63130, United States.
E-mail address: rtreiman@wustl.edu (R. Treiman).

https://doi.org/10.1016/j.cognition.2018.08.016
Received 17 July 2017; Received in revised form 24 August 2018; Accepted 27 August 2018

which they occur in words of their language (Kessler, Pollo, Treiman, & Cardoso-Martins, 2013, for Brazilian prephonological spellers; Pollo, Kessler, & Treiman, 2009, for US and Brazilian prephonological spellers). The productions of prephonological spellers are also influenced by the characteristics of their given name, a word that is particularly important to them. These children use letters from their names rather often when asked to write other words—more often than expected based on the frequency of these letters in words in general (Treiman, Kessler, Boland, Clocksin, & Chen, 2018, for US prephonological spellers; see also Bloodgood, 1999, for US kindergartners and Both-de Vries & Bus, 2008, for Dutch children).

Older prephonological spellers have had, on average, more exposure to writing than younger prephonological spellers. Based on a statistical-learning view, we might expect to see age differences in the productions of prephonological spellers that reflect these differences in experience. Treiman et al. (2018) found some such differences in a cross-sectional study of US prephonological spellers who ranged in age from 3;2 (years; months) to 5;6. The spellings produced by older prephonological spellers were on average longer than those produced by younger ones, and older prephonological spellers were more likely than younger ones to use digrams in proportion to their frequency of occurrence in books. Repeated-letter digrams, such as ‹ee› in ‹coeec›, and alphabetic-order digrams, such as ‹ab› in ‹abeucy›, were less common among older prephonological spellers than among younger ones. Although older prephonological spellers were more likely than younger ones to use digrams in proportion to their frequency of occurrence in English, older and younger prephonological spellers were influenced to the same extent by the frequency of individual letters. The overall proportion of letters in children's spellings that came from their own names did not vary significantly as a function of age, but the position of the name from which these letters came did vary, such that younger prephonological spellers were especially likely to use the first letter of their name when writing other words. This difference may reflect children's increased knowledge of letters in the name beyond the first letters over the preschool years.

In the present study, we asked whether age differences in the characteristics of prephonological spelling are found in a different language, Portuguese. This is an important question because the orthography studied by Treiman et al. (2018), English, is often described as deep. That is, English has complex and sometimes inconsistent links between spellings and sounds. Portuguese orthography is more shallow. Many studies have shown that learning to spell and read is influenced by the orthographic depth of a writing system (e.g., Serrano et al., 2011; Seymour, Aro, & Erskine, 2003). One might expect that learners of a deep orthography would rely heavily on visual patterns and that the learning of such patterns would be more important for a deep orthography such as English than for a shallower orthography such as Portuguese. On the other hand, orthographic depth might not influence the children of primary interest here: those who do not yet use letters for their sound values. In that case, we might see age-related effects in Brazilian prephonological spellers that are similar to those found among US prephonological spellers.

To address these issues, we combined data from previous studies conducted by our research group in Brazil. We conducted a preliminary analysis using the full set of preschool data to determine whether older preschoolers were less likely than younger ones to be prephonological spellers. In our main analyses, we examined just the prephonological spellers. We hypothesized that the productions of older prephonological spellers would differ in some ways from the productions of younger ones, reflecting the older children's greater experience with Portuguese writing in general and their own written names in particular.

2. Method

2.1. Participants

We used data from preschool children who had participated in studies that had been conducted by our research group in Belo Horizonte, Brazil. Table 1 provides information about the studies and the participants. Of the children whose data were included in the studies listed in the table, we used data from children who attended classes designed for 4- and 5-year-olds in private preschools and who produced at least six spellings. At the time the data were collected, children from middle- and upper-class families generally attended private preschools for several years before they reached the age of mandatory school attendance, 6 years. In the preschool classes that the children in our analyses attended, children are read to, learn to write their names, and learn about the alphabet, but they do not receive extensive literacy instruction. In the preliminary analysis that we report, we included all children who fit the criteria mentioned above, regardless of whether their spelling was prephonological. The first column of data in Table 1 shows, for each study and each time point, the number of such children. The second column of data shows the mean ages of these children. This full data set included data from 313 different children (154 girls). There

Table 1
Information about children contributing data to analyses.

Study	All spellers		Prephonological spellers	
	N	Mean age	*N*	Mean age
Cardoso-Martins,[a] Time 1	61	4;3	33	4;3
Cardoso-Martins,[a] Time 2	70	4;6	32	4;5
Cardoso-Martins,[a] Time 3	73	4;10	19	4;10
Cardoso-Martins,[a] Time 4	75	5;3	13	5;1
Cardoso-Martins,[a] Time 5	75	5;6	3	5;0
Cardoso-Martins,[a] Time 6	76	5;10	2	5;8
Cardoso-Martins et al. (2006), Exp. 2, Time 1	43	5;3	13	5;2
Cardoso-Martins et al. (2006), Exp. 2, Time 2	44	5;10	2	5;6
Pollo, Treiman, and Kessler (2008), Exp. 2[b]	46	5;5	7	5;4
Pollo, et al. (2009)	79	4;11	35	4;7
(Pollo, 2008, Exp. 2): subset of children from Pollo et al. (2009) at Time 2, 1 year later	33	5;5	2	5;6
Treiman et al. (2013), Exp. 2	68	4;11	17	4;5
Treiman et al. (2013) follow-up: subset of children from Exp. 2 of Treiman et al. at Time 2, 3 months later	23	4;10	9	4;6

[a] Cardoso-Martins (2013), Kessler et al. (2013), and Treiman et al. (2013) reported analyses of this longitudinal data set examining different questions than those addressed here.

[b] Includes 16 children whose data were excluded from published report because they did not write the letters of interest for those analyses in a dictation task.

were 766 spelling tests, for a number of the children provided data at more than one time point of a longitudinal study. In our main analyses, we focused on children who were identified as prephonological spellers at a particular testing point. The third and fourth columns of data in Table 1 provide information about the prephonological spellers. The prephonological data set included data from 129 different children (61 girls). There were 187 spelling tests, for some children were prephonological spellers at more than one time point.

2.2. Procedure

In each study, children were asked to write a series of items to the best of their ability. Children wrote by hand on paper. Our data repository[1] includes a list of the set of items for each study. Children in some of the studies were tested longitudinally, and Table 1 provides information about the time points. In some cases, the items for a time point were presented over the course of several days. In most of the experiments, the items were dictated by an experimenter. In Experiment 2 of Treiman, Pollo, Cardoso-Martins, and Kessler (2013) and the follow-up study, the words were portrayed in pictures. Children were first familiarized with the pictures and their intended labels and, during the spelling test, were asked to name each picture and then write the word. The experimenter helped the child to use the intended label if the child did not do so. In some studies, children were asked to identify the letters that they used (Pollo et al., 2009) or any letters that were unclear (Treiman et al., 2013) after producing each spelling. The spelling was scored as containing the letters that the child intended. If a child could not identify any letters, the spelling was dropped from the analyses. In other studies, a researcher who was experienced with children's handwritten productions scored the letters in each spelling. Reliability of this scoring was reported to be adequate at Time 1 of the longitudinal study reported in Kessler et al. (2013) and Treiman et al. (2013, Exp. 2). Productions in which the researcher could not discern any letters were excluded from the analyses. We ignored the few diacritics that children used, such as an accent mark on ‹é›. Distinctions between upper- and lower-case letters were also ignored.

3. Results

The first step in our analyses was to identify children whose spelling was prephonological, in that they did not use phonologically plausible letters more often than expected by chance. After describing the procedure used to identify prephonological spellers, we present an analysis using the full set of data to test whether the likelihood that a child was a prephonological speller was lower if the child was older than if the child was younger. We then report our main analyses, which examined characteristics of the prephonological spellers' productions. The data files and R script for the analyses are available in the above-mentioned data repository.

3.1. Identification of prephonological spellers

To identify prephonological spellers from among the larger group, we scored the spellings produced by each child at each time point for their phonological plausibility using the program Ponto (Kessler, 2009). We used the correspondences in Appendix B of Treiman et al. (2013), adding two correspondences that were needed for the present items (/ʎ/ to ‹lh› and /ɲ/ to ‹nh›). To score the phonological plausibility of each spelling produced by a child, the program attempted all possible alignments obtained by associating the phonemes in the target word from left to right with the phonograms (letters or two-letter groups such as ‹lh› that can spell phonemes) in the child's spelling. An association between a phoneme and a phonogram was assigned 1.4 penalty points if the phonogram never spells that phoneme in Portuguese (a phonologically implausible spelling), 1 penalty point if the phoneme was not associated with any phonogram (an omitted phoneme), 1 penalty point if the phonogram was not aligned with any phoneme (an extraneous spelling), and 0 penalty points if the phonogram spells the phoneme in some word of Portuguese (a phonologically plausible spelling). Insertions of ‹h› were not penalized because this letter is silent in Portuguese. The program summed the penalties for each association in each word and reported the total for each child for each spelling test. When there was more than one way of aligning phonemes to phonograms for a word, the program used the one that received the lowest (i.e., best) score. To obtain an indication of the score that a child would receive by chance, the program randomly rearranged the child's spellings of each word with respect to the target words and scored them again. It repeated this procedure 10,000 times and computed the proportion of rearranged spellings for which the score was at least as good as the original score. To identify children for whom there was no convincing evidence that the spellings reflected the phonemes in the targets, we required that the p value was greater than .20 when the spellings were scored in this manner. Although any p value cutoff is arbitrary, we chose a cutoff that was stricter than .05 in order to be more confident that any apparent resemblances between a child's spellings and phonologically plausible spellings were due to chance. To further increase our confidence in the identification of prephonological spellers, we repeated the analyses based on the first phoneme of the targets' pronunciations and the first letter of the children's spellings, ignoring the other phonemes and letters in the items. We identified a child as a prephonological speller if the p value exceeded .20 according to both the first-letter analysis and the whole-word analysis.

3.2. Prephonological spelling as a function of age

To test the hypothesis that the proportion of children identified as prephonological spellers would be lower among older preschoolers than among younger preschoolers, we conducted a mixed-model logistic regression using data from all 766 spelling tests in the full data set. Age (calculated to two decimal points in years) was the fixed factor. We included random intercepts for the identity of the child, given that many children contributed data at more than one time point. We also included random intercepts for the identity of the item set, allowing us to account for variability across studies and time points of a study in the items that children were asked to spell. This and the other mixed-model analyses were conducted using R version 3.5.0 (R Core Team, 2018) and the packages lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2015). The likelihood that a child was identified as a prephonological speller was lower if the child was older than if the child was younger ($b = -2.70$, $SE = 0.34$, $p < .001$).

3.3. Characteristics of prephonological spelling

We now turn to our main analyses, which examine the productions of the prephonological spellers. We treated age as a continuous variable in the analyses, but the tables show the results for prephonological spellers by terciles. The young group had a mean age of 4;1 ($n = 63$; age range from 3;3 through less than 4;5), the middle group had a mean age of 4;8 ($n = 62$; age range from 4;5 through less than 4;11), and the old group had a mean age of 5;3 ($n = 62$; age range 4;11 to 6;0). Some children who were tested longitudinally were included in more than one age group because they were prephonological spellers at more than one point; we controlled for this in the statistical analyses by including the identity of the child as a random factor.

3.3.1. Length

One basic characteristic of a spelling is the number of letters that it contains. We conducted a mixed-model analysis of spelling length to

[1] https://osf.io/kqzuh/?view_only=3c2f86ffd1dd4bdf823b1facca521ca3.

Table 2
Mean values of spelling measures calculated across children in each age group (Standard Deviations in Parentheses).

Measure	Age group		
	Young (mean age 4;1)	Middle (mean age 4;8)	Old (mean age 5;3)
Length	3.33 (2.37)	4.68 (3.26)	5.31 (2.65)
Monogram correlation	0.37 (0.14)	0.40 (0.10)	0.43 (0.10)
Digram correlation	0.16 (0.07)	0.18 (0.06)	0.20 (0.06)
Proportion of digrams with repeated letters	0.09 (0.29)	0.07 (0.25)	0.05 (0.22)
Proportion of trigrams with repeated letters	0.04 (0.19)	0.02 (0.14)	0.01 (0.12)
Proportion of digrams with letters in alphabetic order, of digrams that contain different letters	0.05 (0.22)	0.05 (0.21)	0.05 (0.22)

determine whether older prephonological spellers produced longer spellings than younger ones, as found for US prephonological spellers (Treiman et al., 2018). The model also included the order in which an item was presented in an experimental session, allowing us to ask whether the length of children's spellings changed over the course of a session. Spelling length was log transformed to make the distribution more normal. The fixed factors were the child's age at test, the order of the item on the test day, and the interaction between age and order. Order was log transformed in this and other models that included this variable to improve the normality of the distribution. Continuous variables were centered in this and other models that included interaction terms. Random intercepts were included in the model for the identity of the child and the item set. There was a main effect of age, such that spellings produced by older prephonological spellers contained more letters than spellings produced by younger prephonological spellers ($b = 0.36$, $SE = 0.02$, $p < .001$). The top row of data in Table 2, which shows the mean length of spellings for each age group, illustrates this trend. The main effect of order arose because spellings that were produced later in an experimental session tended to be shorter than spellings that were produced earlier ($b = -0.04$, $SE = 0.01$, $p < .001$). The interaction between age and order was not statistically significant.

3.3.2. Children's use of letters from their names

The prephonological spellers often used letters from their names when writing other words, with an average of 56% of their letters coming from their names. Because some of the names were rather long, and because names tend to include letters that are common in the language as a whole, further analyses are needed to determine whether the prephonological spellers used letters from their names more often than expected on the basis of other factors and to determine how any tendency to overuse name letters may vary with the child's age and the position of the letter in the name. We thus calculated the number of times that each child used each letter in each spelling test. We expressed this as a proportion of all letters used by the child in the spelling test. We log transformed these proportions to reduce skew, and we conducted an analysis to predict proportional letter use from name membership (whether the letter appeared in the child's name in the position of interest), position of the letter in the name (1–12), the child's age, and the interactions among these factors. For children with compound names, such as Luiz Felipe, we considered letters from both parts to belong to the name. Letters that appeared in more than one position of a child's name, such as ‹e› for Luiz Felipe, were excluded from the analysis because position in the name cannot be uniquely defined for such letters. The model included random intercepts for child and item set. To control for other factors that may influence children's tendency to use specific letters, including the letters' frequency of occurrence in the language, the model also included a random intercept for letter. There was a significant effect of name membership ($b = 0.007$, $SE = 0.001$, $p < .001$), such that children were more likely to use letters from their names than expected on the basis of other factors. The effect of name membership was qualified by a three-way interaction involving name membership, position of the letter in the name, and age ($b = 0.003$, $SE = 0.001$, $p < .001$) as well as by an interaction between name membership and age ($b = 0.006$, $SE = 0.002$, $p = .002$), an interaction between name membership and position in the name ($b = -0.002$, $SE = 0.0003$, $p < .001$), and a main effect of age ($b = 0.0005$, $SE = 0.0002$, $p = .02$). To help interpret the three-way interaction involving name membership, position of the letter in the name, and age, we conducted follow-up analyses for letters in each position of the name. For letters in the first to third positions of the name, we found a statistically reliable tendency to overuse letters from the name that was weaker in older prephonological spellers than in younger prephonological spellers. The interaction between name membership and age was statistically significant in the analysis involving the first letter of the name and the analysis involving the second letter of the name. The interaction between name membership and age, although not statistically significant for most positions of the name beyond the third letter, was in the direction of stronger effects of own-name membership for older prephonological spellers than for younger prephonological spellers.

3.3.3. Monogram correlations

To assess the degree to which children used letters in proportion to their frequency in the language and whether this varied as a function of age, we computed for each child a version of the *monogram correlation* measure used in previous studies (Kessler et al., 2013; Treiman et al., 2018). To compute this measure, we summed the frequency with which a child used each letter of the alphabet across all of the spellings in each test and calculated the Kendall rank correlation coefficient (τ_b) between these values and the frequency counts of the letters in texts for preschool children, based on the word counts of Pinheiro (1996). To avoid including words with low dispersion across texts, we omitted words that Pinheiro reported as having zero frequency in texts from the first year of primary school. The counts for the preschool corpus ignored case distinctions and diacritics and were weighted by word frequency. A child who has a positive value on the monogram correlation measure tends to use letters in proportion to their frequency of occurrence in words that appear in the environment, whereas a child who has a value of zero uses letters in a way that is not so influenced. Table 2 shows the mean values of the monogram correlation measure for each age group. A mixed-model analysis using the identity of the child and the item set as random factors and age as a fixed factor found a significant effect of age, such that older children tended to have higher monogram correlations than younger children ($b = 0.06$, $SE = 0.02$, $p < .001$).

In the analyses so far, we computed letter frequencies in Portuguese based on letters' occurrence in any position of a word. We conducted another analysis to examine whether children were sensitive to position in their letter use. Given the salience of letters in the early positions of words, as shown for example by the finding that children were more likely to overuse letters that appeared early in their names than letters that appeared later, we tested for a more general sensitivity to position

by asking whether children tended to begin their spellings with letters that are common specifically in the initial positions of Portuguese words. For example, ‹c› is the most common word-initial letter in our corpus of preschool texts, even though it is not the most common letter across all positions. The letter ‹i›, in contrast, is not very common in word-initial position. If children are sensitive to how often letters appear at the beginnings of words, their use of monograms in the initial position of their spellings might be more closely associated with the frequency of the monograms in the initial positions of words in the corpus of preschool texts than with the frequency of the monograms in other positions. We thus calculated two monogram correlation values for each child: (1) the monogram correlation involving letters in the initial position of the child's spellings and letters in the initial position of the corpus, and (2) the monogram correlation involving letters in the initial position of the child's spellings and in non-initial position in the corpus. A mixed-model analysis with child and item set as random factors and age, position of the letter in the corpus (coded as 1 for initial and 0 for non-initial), and their interaction as fixed effects did not find the significant positive effect of position that would be anticipated if children tended to begin their spellings with letters such as ‹c› that are common in the initial positions of Portuguese words but not necessarily so common in other positions. In fact, there was a significant effect in the opposite direction ($b = -0.15$, $SE = 0.01$, $p < .001$), perhaps because the non-initial counts from the preschool corpus better reflected children's perceptions of overall letter frequency.

3.3.4. Use of digrams as a function of their frequency in the language

To determine whether children used digrams in proportion to their frequency in the book corpus and whether the tendency to do so varied with age, we calculated a *digram correlation* measure for each child at each time point that was analogous to the monogram correlation measure discussed earlier. Specifically, we summed the frequency with which a child used each digram across all of the child's spellings on a test and calculated the Kendall rank correlation coefficient between these values and the frequency counts of the digrams. Digram counts were calculated as for monogram counts, except that we counted the frequencies of immediately adjacent pairs of letters in spellings that have more than one letter. Data from 20 spelling tests in which children produced no two-letter spellings were excluded from the analysis. A positive digram correlation means that a child uses digrams that occur frequently in Portuguese more often than digrams that occur less frequently. Table 2 shows the mean values of the digram correlation measure for each age group. A mixed-model analysis using the identity of the child and the item set as random factors and age as a fixed factor found a significant effect of age ($b = 0.03$, $SE = 0.01$, $p < .001$), such that older prephonological spellers had higher digram correlations than younger ones.

To determine whether children showed a tendency to begin their spellings with digrams that are frequent specifically in the initial positions of Portuguese words, we conducted an analysis similar to that reported for monograms. Children's use of digrams in initial position of their spellings did not correlate significantly more highly with the frequency of the digrams in initial position of the corpus of preschool texts than with the frequency of the digrams in non-initial position. As for monograms, there was a significant effect in the reverse direction ($b = -0.008$, $SE = 0.003$, $p = .017$).

3.3.5. Sequences of repeated letters

Our next analyses examined a specific type of letter sequence: those in which the letters are the same. These sequences are of interest because they are uncommon in Portuguese, with repeated-letter digrams constituting 1% of the digrams in children's texts and repeated-letter trigrams virtually nonexistent. We asked whether the productions of older prephonological spellers reflected these facts better than did the productions of younger prephonological spellers.

The fourth row of data in Table 2 shows the proportion of digrams in children's spellings in which the two letters were the same. To analyze the effects of age and other factors on use of repeated-letter digrams, we coded each digram in each spelling for whether the letters were identical. We then conducted a mixed-model logistic regression analysis to predict whether the letters in a digram were identical. The model included random factors for child, item set, and spelling length and fixed factors for the child's age at test, the order of the item in the session, and the position of the digram in the spelling. In ‹gcaogoo›, for example, ‹oo› was the sixth digram. Interactions of age with both order in the session and position in the spelling were included in the model. There was a main effect of age, such that a digram was less likely to consist of repeated letters if the speller was older than if the speller was younger ($b = -0.82$, $SE = 0.17$, $p < .001$). The main effect of order in the session was also significant ($b = 0.81$, $SE = 0.14$, $p < .001$). This effect arose because children were more likely to use repeated-letter digrams on later trials of a session than on earlier trials. The only other significant effect was the main effect of the position of the digram in the spelling ($b = 0.40$, $SE = 0.15$, $p = .010$). Repeated-letter digrams were more likely to occur in later positions of a spelling, as in ‹gcaogoo›, than in earlier positions of a spelling, as in ‹ooarpfe›.

As shown in the fifth row of data in Table 2, the prephonological spellers occasionally used repeated-letter trigrams. We analyzed the data for repeated-letter trigrams in a way that was analogous to that done for repeated-letter digrams. There was a main effect of age, such that repeated-letter trigrams were less common in older prephonological spellers than in younger ones ($b = -1.35$, $SE = 0.39$, $p < .001$), and a main effect of order, such that repeated-letter trigrams were significantly more likely to occur later in a session than earlier in a session ($b = 2.58$, $SE = 0.47$, $p < .001$). Order interacted with age ($b = -2.24$, $SE = 0.87$, $p = .011$), such that older prephonological spellers showed a smaller increase in the rate of repeated-letter trigrams across a session than did younger prephonological spellers. There was also a main effect of position of the trigram in the spelling ($b = 0.92$, $SE = 0.39$, $p = .017$). Repeated-letter trigrams were more likely to occur in later positions of a spelling than in earlier positions.

3.3.6. Sequences of letters in alphabetic order

Written words in Portuguese do not usually contain pairs of letters in alphabetic order. Of the digrams in children's books, 4% consist of letters in alphabetic order. However, children have the opportunity to learn about the order of the alphabet through such experiences as recitation of the alphabet and exposure to alphabet books, and we might expect them to use some alphabetic-order digrams for this reason. The last row of data in Table 2 shows the proportion of non-repeated-letter digrams in the children's spellings in which the letters were in alphabetic order. When we analyzed these data in the same way as for repeated-letter spellings, we did not find significant effects involving age. However, there was a significant effect of position of the digram in the spelling, such that children were more likely to use alphabetic-order digrams near the end of a spelling than near the beginning ($b = 0.37$, $SE = 0.16$, $p = .020$).

4. Discussion

Statistical learning has often been studied using laboratory tasks involving massed presentation of specially constructed items. For example, participants may see a stream of items in an artificial script during the exposure phase of an experiment, and their knowledge of the orthographic patterns in the items may be assessed later (e.g., Chetail, 2017). In the present study, we examined statistical learning as it takes place in the wild: learning about the letter patterns in the words of a real language. Rather than studying the course of learning in adults or primary-school children who are exposed to novel items in a laboratory (Chetail, 2017; Samara & Caravolas, 2014), we studied the outcome of learning in preschool children who are exposed to written language in

their daily lives. Our analyses focused on preschoolers who, when trying to write words, did not yet use letters for their sound values. Any knowledge that such prephonological spellers show about the letter patterns of their language must reflect knowledge of its visual patterns. Our main goal was to test the hypothesis that older prephonological spellers, who have had on average more exposure to their written language, produce more wordlike spellings than younger prephonological spellers.

We found a number of differences between the productions of older and younger prephonological spellers. Specifically, older prephonological spellers produced longer spellings than younger ones, and they were more likely than younger ones to use individual letters and digrams in proportion to their frequency of occurrence in Portuguese. Older prephonological spellers were less likely than younger ones to produce strings of repeated letters, both repeated-letter digrams and repeated-letter trigrams. In all of these respects, the spellings of older prephonological spellers looked more similar to Portuguese words than did the spellings of younger prephonological spellers. We also found age-related differences in children's use of letters from their own names. The prephonological spellers often used letters from their names when writing other words, but the younger ones differed from the older ones in their greater tendency to overuse letters from the early positions of their names.

In many respects, the present findings with learners of Portuguese agree with findings with learners of other languages. The Brazilian prephonological spellers studied here were similar to the U.S. prephonological spellers studied by Treiman et al. (2018) in showing age effects on spelling length, tendency to use digrams in proportion of their frequency of occurrence in the language, and tendency to repeat letters. The Brazilian prephonological spellers were also similar to the Dutch children studied by Both-de Vries and Bus (2008) and the U.S. children studied by Treiman et al. in that they often used letters from their names when writing other words and in that letters from early positions of the name were especially likely to be overused if children were younger. One difference between the present results and those of Treiman et al. is that the present study found a significant increase with age in prephonological spellers' tendency to use letters in proportion to their frequency of occurrence in the language. Treiman et al. observed a trend in this direction, but it was not statistically significant. Another difference is that alphabet-order sequences were significantly more common in the spellings of younger prephonological spellers than in the spellings of older ones in the study of Treiman et al. while we did not find such a difference in the present study.

Although there were a few differences, the many similarities between the present results with Brazilian prephonological spellers and the previous findings with US prephonological spellers (Treiman et al., 2018) are striking because of the differences in orthographic depth between Portuguese and English. Many studies have shown that the orthographic depth of a writing system strongly influences spelling and reading development (e.g., Serrano et al., 2011; Seymour et al., 2003). One might have anticipated that the learning of visual patterns would play a larger role for learners of English than for learners of Portuguese because of the complexity of sound–spelling relationships in English. The fact that we found few cross-linguistic differences suggests that the complexity of letter–sound associations does not matter before children have begun to use letters to symbolize sounds. The written words of a language can be conceptualized as patterns of letters rather than as representations of language, and early learning of graphotactic patterns seems to occur in similar ways for children exposed to a shallower alphabetic writing system, Portuguese, and children exposed to a deep alphabetic writing system, English. Indeed, some findings suggest that learning of visual patterns occurs in similar ways in Chinese, which differs substantially from alphabetic writing systems in how it represents language (e.g., Luo, Chen, Deacon, & Li, 2011; Qian, Song, Zhao, & Bi, 2015; Tong & McBride-Chang, 2014).

The results of the present study are broadly similar to those of studies that have used artificial materials to examine the learning of letter patterns (e.g., Chetail, 2017; Samara & Caravolas, 2014) in that participants in both types of studies learn about graphotactic patterns without explicit teaching. Studies using artificial materials have often focused on patterns involving position in a string, as when certain letters or digrams only occur at the beginnings of items and others occur only at the ends. Given that participants in previous studies learned these patterns (Chetail, 2017; Samara & Caravolas, 2014), why did the prephonological spellers studied here and by Treiman et al. (2018) not show a sensitivity to position in their use of monograms and digrams? One reason may be that position effects in Portuguese and English are much more subtle than those in the experiments with artificial materials. For example, ‹c› is more likely to occur in initial positions of Portuguese words than in other positions and ‹e› shows the opposite pattern, but both letters occur rather often in both positions. Another reason is that the participants in the above-mentioned studies with artificial materials were adults or children who had already received several years of formal literacy tuition. The children tested here were preschoolers. Older learners of alphabetic writing systems are sensitive to untaught graphotactic patterns that involve position in a word (e.g., Pacton, Perruchet, Fayol, & Cleeremans, 2001; Treiman, 1993), but that sensitivity may not emerge until after the early point in development investigated here.

In many previous studies of statistical learning of graphotactic patterns, both studies with artificial and natural materials, participants' knowledge of patterns was tested by asking them to make explicit judgments about the degree to which previously unseen items fit learned patterns. For example, participants were asked which of ‹pess› or ‹ppes› looks more like an English word (Cassar & Treiman, 1997) or whether a novel letter string seemed to fit the rules that were used to generate items in an earlier phase of the experiment (Samara & Caravolas, 2014). The present study used a more natural task: writing words. This task is often used in educational and research settings to assess children's knowledge about spelling and sound–letter correspondences (e.g., Lee & Al Otaiba, 2017; Ouellette & Sénéchal, 2017). Our results show age-related differences in performance on this production task even among children who do not yet demonstrate knowledge of how letters represent sounds.

Further research will be needed to better understand how and why the characteristics of prephonological spelling change with experience. Consider, for example, the higher proportion of repeated-letter spellings in the productions of younger prephonological spellers than in the productions of older prephonological spellers. This difference may reflect, in part, knowledge of a smaller set of letters in younger prephonological spellers. Also, even if a younger prephonological speller knows the same number of letters as an older one, the tendency to reuse letters and letter sequences that were used on previous trials (Treiman, Decker, Kessler, & Pollo, 2015) may be stronger in younger children than in older ones. A complete model of prephonological spelling would include parameters that reflect the tendency to reuse letters, knowledge about the frequency of letters in the language, knowledge about the frequency of digrams in the language, and other factors. Studies could examine how these parameters vary with the child's age, the order of an item in an experimental session, and the order of a letter in a spelling. The present results suggest that all of these factors are influential. The effects of order in an experimental session and order of a letter in a spelling have not received much attention, and it will be important to investigate their effects further.

Most theories of literacy development focus on the changes that take place after children have begun to use phonology to read and spell words. For example, researchers have examined how children move from spellings that represent only some of the sounds in words to spellings that represent all of the sounds (e.g., Cardoso-Martins, Corrêa, Lemos, & Napoleão, 2006; Ehri, 2015; Gentry, 1982). No meaningful improvements occur during the prephonological period, according to such theories; all prephonological spellers are alike in that they do not

use letters to represent sounds. Our results suggest, to the contrary, that children's spellings become more similar in appearance to the words of their language even before the children start to use letters for their sound values. That is, the spellings improve in some ways. The findings support a statistical-learning view of spelling (Mano, 2016; Treiman & Kessler, 2014) by suggesting that older prephonological spellers, who on the average have had more exposure to writing than younger ones, have benefited from this exposure. Our findings also point to a dissociation between the visual skills that are involved in learning about the appearance of writing and the phonological skills that are required to learn about the links between units of writing and units of language. The older prephonological spellers in our study were somewhat slow to develop the latter skills, for they were prephonological spellers at an age when many of their peers had begun to incorporate phonology into their spelling. What is striking is that the older prephonological spellers had reasonably good visual statistical learning skills and that they applied these skills to the writing that they encountered. They seem to have learned some things about the visual appearance of writing, even though they had not yet begun to use letters for their sound values. Had the older prephonological spellers been poor across-the-board learners, such a result would not have been anticipated.

Of course, the age of a child is only a rough measure of the child's opportunities to learn about the visual patterns of writing. Children of the same age differ from one another in the amount of exposure they have had to written language, the attention that they devote to writing, and the speed at which they learn from exposure. Measuring the characteristics of children's early spellings has the potential to show which children are advantaged in these ways. There is some evidence that these advantages are consequential. For example, Kessler et al. (2013) found that Brazilian prephonological spellers' tendency to use digrams in proportion to their frequency in Portuguese correlated positively with their performance on a standardized spelling test two and a half years later, controlling for age at the preschool test. Children's tendency to use letters from their own names had a negative relationship with later performance. Studies of the statistical-learning skills that enable young children to learn about the graphotactic patterns of their language thus have both practical and theoretical importance.

References

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

Bloodgood, J. W. (1999). What's in a name? Children's name writing and literacy acquisition. *Reading Research Quarterly, 34*, 342–367. https://doi.org/10.1598/RRQ.34.3.5.

Both-de Vries, A. C., & Bus, A. G. (2008). Name writing: A first step to phonetic writing? *Literacy Teaching and Learning, 12*, 37–55.

Cardoso-Martins, C. (2013). Existe um estágio silábico no desenvolvimento da escrita em português? Evidência de três estudos longitudinais [Is there a syllabic stage in spelling development in Portuguese?: Evidence from three longitudinal studies]. In M. R. Maluf, & C. Cardoso-Martins (Eds.). *Alfabetização no século XXI* (pp. 82–108). Porto Alegre: Penso Editora.

Cardoso-Martins, C., Corrêa, M. F., Lemos, S., & Napoleão, R. F. (2006). Is there a syllabic stage in spelling development? Evidence from Portuguese-speaking children. *Journal of Educational Psychology, 98*, 628–641. https://doi.org/10.1037/0022-0663.98.3.628.

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology, 89*, 631–644. https://doi.org/10.1037//0022-0663.89.4.631.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief audience experience. *Cognition, 87*, B69–B77. https://doi.org/10.1016/S0.

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition, 163*, 103–120. https://doi.org/10.1016/j.cognition.2017.02.015.

Ehri, L. C. (2015). How children learn to read words. In A. Pollatsek, & R. Treiman (Eds.). *Oxford handbook of reading* (pp. 293–310). New York, NY: Oxford University Press.

Gentry, J. R. (1982). An analysis of developmental spelling in "GNYS AT WRK". *The Reading Teacher, 36*, 192–200.

Kessler, B. (2009). Ponto [Computer software]. Available at http://spell.psychology.wustl.edu/ponto.

Kessler, B., Pollo, T. C., Treiman, R., & Cardoso-Martins, C. (2013). Frequency analyses of prephonological spellings as predictors of success in conventional spelling. *Journal of Learning Disabilities, 46*, 252–259. https://doi.org/10.1177/0022219412449440.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests in linear mixed effects models. R package version 2.0-29. Retrieved from < http://cran.r-project.org/package=lmerTest > .

Lee, J. A. C., & Al Otaiba, S. (2017). End-of-kindergarten spelling outcomes: How can spelling error analysis data inform beginning reading instruction? *Reading & Writing Quarterly, 33*, 226–238. https://doi.org/10.1080/10573569.2016.1165639.

Luo, Y. C., Chen, X., Deacon, S. H., & Li, H. (2011). Development of Chinese orthographic processing: A cross-cultural perspective. *Writing Systems Research, 3*, 69–86. https://doi.org/10.1093/wsr/wsr008.

Mano, Q. R. (2016). Developing sensitivity to subword combinatorial orthographic regularity (SCORe): A two-process framework. *Scientific Studies of Reading, 20*, 231–247. https://doi.org/10.1080/10888438.2016.1141210.

Ouellette, G., & Sénéchal, M. (2017). Invented spelling in kindergarten as a predictor of reading and spelling in Grade 1: A new pathway to literacy, or just the same road, less known? *Developmental Psychology, 53*, 77–88. https://doi.org/https://doi.org/10.1037/dev0000179.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130*, 401–426. https://doi.org/10.1037//0096-3445.130.3.401.

Pinheiro, Â. M. V. (1996). *Contagem de freqüência de ocorrência de palavras expostas a crianças na faixa pré-escolar e séries iniciais do 1º grau [Frequency count of words children are exposed to in preschool and the first years of primary education] [Computer software].* São Paulo, Brazil: Associação Brasileira de Dislexia.

Pollo, T. C. (2008). *The nature of young children's phonological and nonphonological spellings.* St. Louis, MO: Washington University (Dissertation).

Pollo, T. C., Kessler, B., & Treiman, R. (2009). Statistical patterns in children's early writing. *Journal of Experimental Child Psychology, 104*, 410–426. https://doi.org/10.1016/j.jecp.2009.07.003.

Pollo, T. C., Treiman, R., & Kessler, B. (2008). Preschoolers use partial letter names to select spellings: Evidence from Portuguese. *Applied Psycholinguistics, 29*, 195–212. https://doi.org/10.1017/S0142716407080095.

Qian, Y., Song, Y.-W., Zhao, J., & Bi, H.-Y. (2015). The developmental trend of orthographic awareness in Chinese preschoolers. *Reading and Writing: An Interdisciplinary Journal, 28*, 571–586. https://doi.org/10.1007/s11145-014-9538-8.

R Core Team (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by eight-month-old infants. *Science, 274*, 1926–1928. https://doi.org/10.1126/science.274.5294.1926.

Samara, A., & Caravolas, M. (2014). Statistical learning of novel graphotactic constraints in children and adults. *Journal of Experimental Child Psychology, 121*, 137–155. https://doi.org/10.1016/j.jecp.2013.11.009.

Serrano, F., Genard, N., Sucena, A., Defior, S., Alegría, J., Mousty, P., ... Seymour, P. H. K. (2011). Variations in reading and spelling acquisition in Portuguese, French and Spanish: A cross-linguistic comparison. *Journal of Portuguese Linguistics, 9*, 183–204. https://doi.org/10.5334/jpl.106.

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

Tong, X., & McBride-Chang, C. (2014). Chinese children's statistical learning of orthographic regularities: Positional constraints and character structure. *Scientific Studies of Reading, 18*, 291–308. https://doi.org/10.1080/10888438.2014.884098.

Treiman, R. (1993). *Beginning to spell: A study of first-grade children.* New York, NY: Oxford University Press.

Treiman, R., Decker, K., Kessler, B., & Pollo, T. C. (2015). Variation and repetition in the spelling of young children. *Journal of Experimental Child Psychology, 132*, 99–110. https://doi.org/10.1016/j.jecp.2014.12.008.

Treiman, R., & Kessler, B. (2014). *How children learn to write words.* New York, NY: Oxford University Press.

Treiman, R., Kessler, B., Boland, K., Clocksin, H., & Chen, Z. (2018). Statistical learning and spelling: Older prephonological spellers produce more wordlike spellings than younger prephonological spellers. *Child Development, 89*, e431–e443. https://doi.org/10.1111/cdev.12893.

Treiman, R., Pollo, T. C., Cardoso-Martins, C., & Kessler, B. (2013). Do young children spell words syllabically? Evidence from learners of Brazilian Portuguese. *Journal of Experimental Child Psychology, 116*, 873–890. https://doi.org/10.1016/j.jecp.2013.08.002.