



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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



Statistical patterns in children's early writing

Tatiana Cury Pollo *, Brett Kessler, Rebecca Treiman

Department of Psychology, Washington University in St. Louis, St. Louis, MO 63130, USA

ARTICLE INFO

Article history:

Received 19 February 2009

Revised 7 July 2009

Available online 18 August 2009

Keywords:

Prephonological

Spelling

Print exposure

Statistical learning

Cross-linguistic

Portuguese

ABSTRACT

Many theories of spelling development claim that before children begin to spell phonologically, their spellings are random strings of letters. We evaluated this idea by testing young children (mean age = 4 years 9 months) in Brazil and the United States and selecting a group of prephonological spellers. The spellings of this prephonological group showed a number of patterns that reflected things such as the frequencies of letters and bigrams in children's language. The prephonological spellers in the two countries produced spellings that differed in some respects, consistent with their exposure to different written languages. We found no evidence for reportedly universal patterns in early spelling such as the idea that children write one letter for each syllable. Overall, our results reveal that early spellings that are not phonological are by no means random or universal and preserve certain patterns in the writing to which children have been exposed.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Young children often attempt to write words and sentences after they have learned how to write letters of the alphabet but before they learn how letters represent sounds. Ehri (1991) reported on a child writing *hs* for *quick*, and Bissex (1980) related how a 4-year-old boy wrote a banner with the letters *sshidca* to tell his mother *welcome home*. Even for literacy researchers who are experienced in deciphering children's spelling errors, productions such as these appear to be hopelessly opaque. They appear to have no particular visual connection to the way adults write these words, nor is there evidence that the children have applied any knowledge of how different letters represent specific sounds of language. The goal of this study was to understand the nature of those early, prephonological, spellings. Are they the random concatenations of letters, as they appear to be, or do they reflect some understanding of the structure of written text on the part of the children?

* Corresponding author.

E-mail address: tpollo@wustl.edu (T.C. Pollo).

Most theories of early literacy acquisition concentrate on how children learn to map sounds to phonetically appropriate letters (Ehri, 2005; Gough & Hillinger, 1980). Researchers examine how children analyze words in spoken language as strings of phonemes and how they grasp the idea that letters in words in written language represent phonemes in spoken language (e.g., Liberman, Shankweiler, Fischer, & Carter, 1974). This phonological perspective focuses on phonological development, and it gives short shrift to very early spellings. Before children learn how letters correspond to sounds, their writing is often characterized as random (Gentry, 1982).

Other researchers do study earlier attempts at spelling. They believe that prephonological spellings have patterns that reflect hypotheses that children construct, guided by principles that hold across languages. This perspective is especially well represented in many countries where languages other than English are spoken, including Spanish (e.g., Ferreiro & Teberosky, 1982) and Portuguese (e.g., Martins & Silva, 2001; Rego, 1999; Silva & Alves-Martins, 2002). Similar viewpoints are represented in the emergent literacy tradition in the United States (e.g., Sulzby, 1985). We refer to these researchers as having a constructivist perspective because work in this tradition has been influenced by Piaget's theory and methodology for studying how children construct a view of the world. Ferreiro and colleagues were particularly influential in applying the Piagetian framework to literacy development (Ferreiro, 1990; Ferreiro & Teberosky, 1982; Vernon & Ferreiro, 1999).

Constructivists propose that children know a good deal about writing before they understand that letters encode phonemes and that this knowledge is reflected in their own writing. Among the patterns that constructivists have proposed are the following:

- *Minimum quantity*: Children think that a text needs to have several letters. For example, children are more likely to accept the sequence *bdc* as a word than to accept the sequence *bd* as a word.
- *Within-word variation*: The letters in a word must be different from each other. For example, children are more likely to prefer *bdc* over *bbb*.
- *Between-word variation*: Different words should be written differently. For example, if a child spelled *cat* as *abc*, then *dog* should not be spelled as *abc*. Children may spell different words by arranging the same letters differently. If a child spelled *cat* as *abc*, then *dog* may subsequently be spelled as *bca* or *cba*.
- *Syllabic spelling*: Each letter is believed to represent an entire syllable. For example, Ferreiro & Teberosky (1982) reported that one child spelled the three-syllable Spanish word *patito* 'duckling' using three symbols that look roughly like *cuo*.

Advocates of constructivism claim that these principles are relatively abstract and universal, being formed in similar ways by children learning a variety of languages and scripts. For example, Ferreiro, Pontecorvo, & Zuccheromaglio (1996) suggested that children's preference for variation is independent of the frequency of doubled letters in the writing to which they are exposed. The relatively abstract nature of the principles is shown by children's preference for the pseudoword *bdc* over the too-short *bd* or the too-homogeneous *bbb* even when children have never seen the sequence *bdc*.

Much of the evidence for the constructivists' patterns comes from case studies, anecdotes, and clinical interviews that have not been conducive to rigorous experimental design and statistical analysis. In addition, there is a lack of evidence for the syllabic stage—perhaps the most distinctive aspect of the theory—in certain languages. Kamii, Long, Manning, & Manning (1990), for example, did not find evidence for a syllabic stage among English-speaking children. And a few studies have recently questioned the existence of the syllabic stage even in Portuguese (Cardoso-Martins, Corrêa, Lemos, & Napoleão, 2006; Pollo, Kessler, & Treiman, 2005), where syllabic spellings had been reported previously (Nunes Carraher & Rego, 1984; Rego, 1999).

As we have described so far, most researchers see early spellings either as random (the phonologically oriented tradition) or as having patterns that are guided by universal principles (the constructivist tradition). The hypothesis we pursued in this study is that children's early spellings do have patterns but that these patterns are not universal; they instead reflect statistical patterns that children observe in the texts they see. We call our perspective the statistical learning view.

Support for the general statistical learning perspective comes from evidence that the letter patterns that people see in their daily lives influence their reading and spelling (e.g., Thompson, Cottrell, &

Fletcher-Flinn, 1996). Children and adults appear to acquire regularities by attending to co-occurring patterns and frequencies in words (see Deacon, Conrad, & Pacton, 2008, for a review). For example, Treiman, Kessler, & Evans (2007) showed that adults have a stronger tendency to pronounce word-initial *g* as /dʒ/ as in *gentle* (phonemes are represented using the alphabet of the International Phonetic Association, 1999) in pseudowords that have a latinate suffix like *-ic* than in other types of pseudowords, apparently a statistical generalization from encountering words such as *geriatric* and *generic*. As another example, young children overuse letters of their own names when trying to write other words (Treiman, Kessler, & Bourassa, 2001). According to Bloodgood (1999), this is true even for children who do not yet connect sounds to letters. Her data suggest that such children's letter choices when writing words are neither random nor universal; on average, 41% of the letters they wrote came from their own names. Children apparently overuse those letters because of the disproportionate frequency with which they attend to the spellings of their own names.

If the statistical learning evinced by children with regard to their own names extends to other text in their environment, as has been found for adults, it could show up in their early attempts at writing even before they learn how letters represent sounds. After all, children in most modern cultures have frequent opportunities to see letters in contexts such as picture books, labels on consumer goods, and street signs. It is possible that children pick up certain formal or graphic properties of writing, such as the frequency with which different letters are used and juxtaposed, before they have had the opportunity to learn the functional properties, notably the encoding of phonemes. If the statistical learning perspective is correct in holding that such learning is plausible, a young child's writings may reflect statistical properties of the text the child sees. Moreover, differences in children's textual environments should be reflected in differences in their writing. *Mary* will write a little differently from *John* if these children have typical experiences with their own written names. But if they grow up surrounded by English text, their productions will be more similar to each other than to those of *Luiz*, who grows up surrounded by Portuguese.

As statistical learning adherents, we agree with constructivists in expecting there to be discernible patterns in prephonological writing. However, although constructivism emphasizes constructions that are universal, we emphasize that children's writings reflect their input. We expect to find differences among children who speak different languages because the children have been exposed to different textual input. We analyzed the writings of young children in the United States and Brazil to find out whether the so-called random letter productions of young children preserve certain characteristics of texts written in the children's languages. There has previously been no replicable way to decide whether an individual child is prephonological. Therefore, we developed a statistical procedure for testing whether a child has any tendency to use letter sounds (or letter names) when spelling words. This allowed us to identify and contrast groups of children who are either clearly phonological or clearly prephonological in their spelling productions. Having ruled out the possibility that prephonological children are influenced by conventional functional considerations (phoneme encoding), we looked for other patterns in their spellings. We investigated the patterns that constructivists have proposed and the patterns we would expect if children's spellings reflect statistical properties of written text. By looking for differences between English-speaking children from the United States and Portuguese-speaking children from Brazil, we sought to distinguish between the constructivist position that patterns tend to be universal and the statistical learning position that patterns tend to reflect children's textual environment.

Three sources of information about the textual environment were exploited. The first source was reading materials targeted to young children in the respective countries; with them, we compared the frequency distribution of the different letters and of their juxtapositions (bigrams) and the use of consonant and vowel letters, including their frequency and patterns of alternation. The second source was the spelling of the individual children's own names; here we tested with provably prephonological writers Bloodgood's (1999) idea that such children overuse letters from their own names. A final source of information was alphabetical order, familiar to even young children through recitation, alphabet songs, and educational materials such as alphabet strips; here we tested whether children tended to write letters in alphabetical order when spelling words. Children learn about some of these things orally as well as through exposure to writing, and they learn about them explicitly as well as implicitly. Our interest is in whether children's exposure to this information influences their performance in a context outside of that in which it is learned—the production of written spellings.

2. Method

2.1. Participants

Middle- and upper middle-class children were recruited from private preschools during the middle of the school year in Belo Horizonte, Brazil, and St. Louis, Missouri, USA. Among these, three Brazilian children and six U.S. children were excluded from the study because they often used symbols other than letters in their spellings. Children who used a number or another symbol only once or twice were not excluded. The Brazilian group consisted of 79 native speakers of Portuguese whose ages ranged from 3 years 10 months to 6 years 0 months (mean age = 4 years 10 months). The U.S. group consisted of 51 native speakers of English whose ages ranged from 3 years 7 months to 5 years 6 months (mean age = 4 years 8 months).

2.2. Stimuli

Each group of children spelled 18 words and 18 nonwords, as listed in [Appendix A](#). The words and nonwords were evenly distributed into three types of patterns of consonant (C) and vowel (V) sequences: CVC, CCV, and CVCV with stress on the first syllable. The stimuli contained no whole consonant letter names such as /bi/ (English *b*). Vowels that can constitute letter names, such as /e/ for English speakers (*a*) and /ɛ/ for Portuguese speakers (the name of *e*), were balanced between languages for each stimulus pattern. All of the words were expected to be familiar to children but not among the words they would be able to spell at the beginning level.

2.3. Procedure

The children's main task, the spelling task, was to spell the 36 stimuli. The children were also given three tests to evaluate their general literacy level. The tasks were administered over three sessions that were approximately 1 week apart. Each session consisted of one third of the spelling task followed by the letter name task in the first session, the letter sound task in the second session, and the reading task in the third session. All tasks were administered by a native speaker of the relevant language.

2.4. Letter name task

Children were asked to identify all letters of the alphabet, totaling 26 letters in the United States and 23 letters in Brazil, which uses *k*, *w*, and *y* only in borrowed words or proper names. A board with uppercase colored letters in a random order was placed in front of each child, and the child was asked to choose the letter that corresponded to the name spoken by the experimenter. The letters were queried in a different random order for each child.

2.5. Letter sound task

Using the same board, children were asked to choose the letter that spelled the sound produced by the experimenter. The letter sounds were presented in a different random order for each child. U.S. children were tested on /æ/, /ɑ/, /b/, /d/, /ɛ/, /f/, /g/, /h/, /ɪ/, /k/, /l/, /m/, /n/, /p/, /ɹ/, /s/, /t/, /ʊ/, and /ʌ/, and Brazilian children were tested on /b/, /d/, /e/, /f/, /g/, /h/, /l/, /m/, /o/, /p/, /s/, /t/, and /z/. These were the same sounds used by [Pollo et al. \(2005\)](#).

2.6. Reading task

The experimenter showed children 11 different cards with two words and a picture, one card at a time, and asked each child to identify any items he or she knew. If the child did not identify the items, the experimenter pointed to each item and asked whether the child knew it. The order of presentation

of the cards was randomized for each child, and the experimenter praised every response from the child. Only the reading of the words was scored; the pictures were included to make the task less frustrating for nonreaders. All of the words were printed in uppercase letters and were frequently used in kindergarten books of the respective country. For the English-speaking children, the words were the same ones used by Treiman & Rodriguez (1999): *book, come, dog, eat, go, green, in, is, it, jump, look, no, play, red, see, stop, the, up, yellow, yes, you, and we*. For the Portuguese-speaking participants, the words were similar in difficulty and frequency: *alto* 'high', *amarelo* 'yellow', *azul* 'blue', *bola* 'ball', *chuva* 'rain', *comeu* 'ate', *em* 'in', *eu* 'I', *gato* 'cat', *joga* 'plays', *livro* 'book', *não* 'no', *nós* 'we', *olhe* 'look', *pula* 'jumps', *sou* 'am', *três* 'three', *um* 'one', *vai* 'goes', *vamos* 'let's go', *verde* 'green', and *você* 'you'.

2.7. Spelling task

Half of the children were randomly selected to spell the real words in the first one and a half sessions and to spell the nonwords in the last one and a half sessions; the order was reversed for the other half of the children. The order of the words and nonwords was randomized for each child. The spelling task was presented with the aid of a cat puppet and a dog puppet. One puppet dictated the first 18 stimuli, and the other puppet dictated the following 18 stimuli; thus, one puppet dictated words, and the other puppet dictated nonwords. The experimenter explained that the puppet wanted to see how children spelled words. For the nonword condition, the experimenter added that the puppet liked to say funny words that did not mean anything. The puppet said the word or nonword, used it in a sentence, and repeated it. The children were asked to say the word or nonword before spelling it, and they were told that we were not concerned with the accuracy of their spellings. As each child produced the spellings, we asked him or her to identify each letter that was used. In rare cases, a child's intended letters did not seem to be what the child in fact wrote, and in those circumstances what the child said to be the letter prevailed.

2.8. Textual environment statistics

For analyses comparing Portuguese-speaking children's productions with their textual environment, we used a frequency list of words found in a corpus based on children's reading material used for pedagogical purposes in Belo Horizonte (Pinheiro, 1996), selecting the 3,621 word types (total frequency = 31,889 tokens) that appear in both the preschool and first-grade subcorpora at least once. For English, we used the 6231 words (total frequency = 796,265) that appear in both the kindergarten and first-grade lists of Zeno, Ivenz, Millard, & Duvvuri (1995) at least once. However, books are not the only location of text for children, and the importance of books might not be as great as is often presumed; children sometimes do not even look at the print when they are being read to (Evans & Saint-Aubin, 2005). Another salient type of text for young children is the written forms of their own names and those of their peers (Levin & Ehri, 2009; Share & Gur, 1999). Sometimes parents deviate from the standard spelling patterns of the language when naming their children. Therefore, we performed parallel analyses on a list of 493 Brazilian names (204 different types) and 548 American names (335 different types). These were names of children enrolled in preschools patronized by families of similar socioeconomic status as the children who participated in the study.

3. Results

Table 1 shows data on children's performance on the letter name, letter sound, and reading tasks. Even though the Brazilian group was older on average than the U.S. group, the two groups did not differ significantly on these tasks ($p > .16$ for all). The finding that Brazilian children were slightly older than their U.S. counterparts with similar educational experiences but displayed a similar level of pre-literacy skills is consistent with previous studies (e.g., Treiman, Kessler, & Pollo, 2006).

Table 1

Mean proportions of correct responses in the letter name, letter sound, and reading tasks in Portuguese and English for all children and for the prephonological and phonological spellers.

Task	All children		Prephonological spellers		Phonological spellers	
	Portuguese (<i>n</i> = 79)	English (<i>n</i> = 51)	Portuguese (<i>n</i> = 35)	English (<i>n</i> = 23)	Portuguese (<i>n</i> = 31)	English (<i>n</i> = 21)
Letter name	.81 (.26)	.82 (.21)	.70 (.26)	.76 (.22)	.99 (.04)	.92 (.12)
Letter sound	.41 (.31)	.49 (.25)	.24 (.18)	.40 (.21)	.69 (.04)	.63 (.27)
Reading	.09 (.24)	.15 (.27)	.00 (.00)	.06 (.11)	.22 (.34)	.28 (.36)

Note. Standard deviations are in parentheses.

3.1. Random and phonological spellers

Children whose spellings are random from a phonological standpoint were the major focus of this study, and so it was important to determine which children were prephonological spellers. To do so, we generated all phonologically plausible spellings for each item, using not only the orthographically correct letter in the case of words but also letters and digraphs that are often used to spell the sound in other words and those that turn up often in phonological spellers' errors. For example, in Portuguese *h* was accepted as a plausible spelling for /g/ or /ga/ in stimuli such as *gado* 'cattle', reflecting the finding (Pollo, Treiman, & Kessler, 2008) that young children's phonological spellings are influenced by their knowledge of the letter name for *h*, which is /aga/ in Portuguese. Thus, *hdo*, *hdu*, *hado*, and *hadu* all would be accepted as phonological, along with the more obvious *gadu* (final *-o* sounds like /u/ in Portuguese) and the correct *gado*. We assigned a phonological plausibility score to each child's spelling by measuring the string edit, or Levenshtein, distance (Kruskal, 1983) between it and each of the phonologically plausible spellings and using the best of those distances. Our Levenshtein metric counted one unit of distance for each letter addition, deletion, or substitution that is necessary for transforming the attested spelling into a plausible spelling. If the child's spelling matched a real or plausible spelling completely, the spelling got a distance score of 0; the greater the deviation of the spelling from the word being compared, the higher the score. For example, if a child spelled *gado* as *hbug*, the spelling would get a distance measure of 2 because it can be turned into plausible *hdu* by substituting *d* for *b* and by deleting *g*. We summed the scores for all 36 spellings for each child. We then ran Monte Carlo simulations to find the probability that the child would get the same or better scores by chance. We randomly rearranged the child's spellings 10,000 times with respect to their target spellings and counted what fraction of those rearrangements had a score at least as good as the child's attested score. If fewer than .05 of the rearranged scores were as good as or better than the child's score, we accepted the hypothesis that the child was spelling phonologically. We found 31 Portuguese-speaking phonological spellers (mean age = 5 years 4 months) and 21 English-speaking phonological spellers (mean age = 4 years 11 months) among our participants. All of the children whose spelling scores were neither significantly better than chance level nor more than 1% better than the average score of the rearrangements were considered as prephonological spellers. These criteria yielded 35 Portuguese-speaking children (mean age = 4 years 8 months) and 23 English-speaking children (mean age = 4 years 7 months). The following analyses concentrate on this latter nonphonological (or prephonological) group, in some cases comparing them with the former phonological group. Note that some participants (13 Portuguese-speaking and 9 English-speaking children) could not be placed confidently into either group, so their spellings are not analyzed further. Table 1 shows the literacy measures for the prephonological and phonological children.

3.2. Evidence for constructivism

We first investigated whether the prephonological spellers' productions conformed to the main constructivist hypotheses.

3.2.1. Minimum quantity hypothesis

In this analysis, we investigated whether children are reluctant to use fewer than three letters in their spellings. Fig. 1 shows the proportions of spellings of different lengths in children's spellings and in the corpora (the figure aggregates the longer infrequent lengths into one category of more than 10 letters). Three-letter spellings were the most common length for Portuguese-speaking children, and four-letter spellings were the most common length for English-speaking children. Even though three- and four-letter spellings were the most common lengths, one- and two-letter spellings were not avoided nearly as often as the minimum quantity hypothesis would suggest. Portuguese-speaking children used one or two letters in their spellings 22% of the time, and English-speaking children did so 13% of the time. Those numbers did not differ significantly from what is found in running texts of Portuguese and English (27% and 22%, respectively, $p > .19$ for both). In fact, the length distribution in children's spellings correlated significantly with the length distribution in Portuguese words, $\rho = .845$, $p < .001$, and in English words, $\rho = .873$, $p < .001$. Here and throughout the article, we used Spearman's rank correlation coefficients when the distribution of the variable was skewed. The results suggest that children pick up characteristics of written texts and mirror some of them in their own productions.

3.2.2. Within-word variation

We next investigated whether children avoid producing sequences of repeated letters such as *bb*. We concentrated on sequences of two letters, or geminates, because both languages use them, albeit with different frequencies. In Portuguese children's texts, 1% of all two-letter sequences are geminates, and only four different letters can double. In English texts, 4% of all such sequences are geminates, and 18 different letters can double. In our list of Brazilian names, we counted a geminate rate of 2%, as compared with 5% for American names.

We looked at the number of geminates in children's spellings in both languages. The probability that children would produce geminate bigrams by chance was computed by randomly rearranging the letters in each child's spellings and counting the mean number of geminates in the rearranged data. If, as proposed in the constructivist framework, children hesitate to use sequences of the same

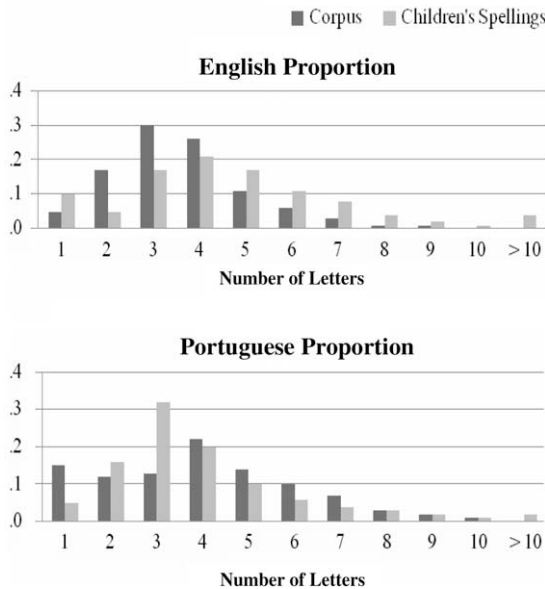


Fig. 1. Proportions of spellings of various lengths in children's prephonological spellings and in the corpora for English and Portuguese.

letters, the number of geminate bigrams in children's scores should be lower than the rearranged scores (chance). According to the constructivist framework, children's preference for variation is independent of the frequency of doubled letters in the writing systems to which they are exposed. We observed geminate spellings at rates of 13% for English and 4% for Portuguese; the rates expected by chance were 14% and 13%, respectively. These relatively high expectations for doubled letters reflect the fact that prephonological spellers often use a fairly small number of different letters. A by-participant analysis of variance (ANOVA) with type of score (actual or rearranged) as a within-subject factor and language as a between-subject factor showed a significant effect of type of score, $F(1, 56) = 19.06$, $p < .001$, $\eta^2 = .254$ (all η^2 values reported are partial). This effect shows that children were less likely to use geminate bigrams than was expected by chance. The difference between the actual scores and the rearranged scores was statistically reliable in both countries, but the significant interaction between score type and language, $F(1, 56) = 8.65$, $p = .005$, $\eta^2 = .134$, indicated that Portuguese speakers were more likely than English speakers to avoid geminates. We found the same pattern of results when we carried out the analysis without children who had geminates in their names. The observed difference between countries is contrary to the constructivist assumption of universality, but it reflects the differences between English and Portuguese texts. This demonstrates that children tend to preserve some characteristics of the text in their spellings.

3.2.3. Between-word variation

To test the constructivist idea that young children believe that different words must be written differently, we counted the number of times that children wrote different stimuli exactly alike. Because children would be unlikely to remember their invented spellings from one week to the next, we counted the number of repetitions in each day of testing separately. We then randomly rearranged the letters that children used, keeping the number of letters constant for each word, and counted the number of repetitions that occurred in this new set of rearranged words. If children tend not to repeat the same arrangement of letters for different words, they should show less repetition in their own productions than in the rearranged words. We found the opposite result; children in both countries repeated spellings in their observed productions more than in the rearranged spellings. Portuguese-speaking children made 159 repetitions in their own spellings versus only 75 repetitions in the rearranged words. English-speaking children's spellings had 76 repetitions, and the rearranged words had only 55 repetitions. In both languages, not a single one of the random rearrangements of the data had more repetitions than the children's observed data (significance of the one-tailed hypothesis that children repeat less often than chance: $p = 1.0$). That is, children reused the same spellings more often than was expected by chance, the opposite pattern of that predicted by the constructivist approach.

3.2.4. Syllabic spelling

The last constructivist principle that we investigated with our prephonological group is the idea that these children write one symbol per syllable. A strict interpretation of this hypothesis is difficult because it would be expected to interact with the minimality hypothesis, which holds that children resist using fewer than three symbols. Therefore, instead of asking whether children used exactly the same number of symbols as syllables, we adopted the more lenient criterion used by Cardoso-Martins et al. (2006) and asked whether children used more symbols when spelling the disyllabic stimuli (CVCV) than the monosyllabic stimuli (CCV and CVC). Table 2 shows the mean numbers of letters for the prephonological spellers broken down by language and consonant–vowel structure.

An ANOVA with language as a between-subject factor and consonant–vowel structure (CCV, CVC, or CVCV) as a within-subject factor did not reveal any significant effects. The lack of a significant effect of structure, $F(2, 112) = 0.84$, $p = .43$, indicates that prephonological spellers used the same number of letters for all stimuli, regardless of the number of syllables.

3.3. Influence of the textual environment

In this set of analyses, we asked whether children's spellings were similar, in terms of various characteristics, to those of the writing system to which they were exposed. We have already seen a

Table 2

Mean numbers of letters in prephonological spellings in English and Portuguese.

Consonant–vowel structure	English	Portuguese
CVC	4.80 (2.00)	3.98 (1.81)
CVCV	4.94 (2.12)	3.99 (1.83)
CCV	4.88 (2.02)	4.01 (1.97)

Note. Standard deviations are in parentheses.

similarity in terms of the distribution of spelling lengths and use of geminates, and here we examined a number of other characteristics. Our main interest was in whether the productions of the prephonological spellers share some of the characteristics of the writing to which they have been exposed, but for some analyses we include the results of the phonological group for comparative purposes. For each characteristic that we investigated, we first conducted a set of analyses on our lists of words—both from books and from children’s names—to identify differences between texts in the two languages. In general, we report analyses by word token frequency; for example, English has only three one-letter words, a minuscule proportion of the entire vocabulary, but they occur 41,135 times in children’s reading materials, so we count 41,135 one-letter words in English text, more than 5% of all word tokens. Our hypothesis is that prereaders are relatively unlikely to abstract away from token repetitions (e.g., to conceptualize all instances of *a* as just one object type) if they have no idea what words the spellings stand for. Nevertheless, we also computed parallel analyses by type, counting each word only once, and report those analyses here and for the following experiment only when they are substantially different from the by-tokens analyses.

3.3.1. Frequency of letters

Portuguese and English use the same Latin alphabet, but the frequency distributions of those letters differ. For example, *a* is more frequent in Portuguese than in English, whereas the opposite holds for *e*. The frequency distributions for the book corpora and the list of children’s names correlated markedly in both languages, $\rho = .925$ for Portuguese and $\rho = .819$ for English, $p < .001$ for both.

We computed how frequently children used each letter of the alphabet in English and in Portuguese. To analyze the influence of the corpora distribution on children’s spellings, a by-items correlation was performed (26 letters in each language; we included *k*, *w*, and *y* for Portuguese because they are used in some Portuguese words) between each letter’s frequency in the book corpus and in the children’s spellings. There was a significant correlation between frequency distribution in the corpora and in children’s spellings for the prephonological groups, $\rho = .718$ for English and $\rho = .913$ for Portuguese, $p < .001$ for both, as well as the phonological groups, $\rho = .853$ for English and $\rho = .945$ for Portuguese, $p < .001$ for both. The correlation in letter frequencies might be intermediated by previously reported propensities for children to use letters from their own names. To control for that possibility, we performed the same analysis using only children who did not have the given letter in their names. For example, we looked at children’s use of *a* only for those children who did not have *a* in their names. The correlations between the frequency distribution in the corpora and in the spellings remained significant. The correlations for the phonological spellers may be mediated by the fact that they selected the correct letters a substantial proportion of the time and the word stimuli were quite typical in each language. The correlations for the prephonological children are more likely to reflect a direct influence of the letters’ textual frequency because the randomization tests showed no significant correlation with the correct spellings.

3.3.2. Vowels and consonants

Another type of analysis that was conducted at the letter level concerns the distribution of consonant and vowel letters in children’s spellings. This categorization is interesting because of prior reports of all-vowel spellings in languages such as Spanish and Portuguese as opposed to all-consonant spellings in languages such as English (see, e.g., Pollo et al., 2005). Consonant–vowel categorizations might not be a natural classification for young children who do not yet know the sounds of all the letters, but

they provide a way of summarizing trends over a wide range of letters in a way that adults find meaningful and interesting. Because *y* and *w* are ambiguous letters, we counted only *a*, *e*, *i*, *o*, and *u* as vowels.

Pollo et al. (2005) reported that Portuguese has a higher percentage of vowel letters in its words (51%) than does English (39%) and that children in Brazil use more vowels than do children in the United States. In the current study, we computed the percentage of vowel letters in the list of children's names and found that Brazilian names are 49% vowels and American names are 40% vowels. We also computed the percentage of vowels in children's spellings. The results, shown in Table 3, were subject to an ANOVA with language and speller group (prephonological or phonological) as between-participant factors and consonant–vowel structure (CVC, CVCV, or CCV) and lexicality (word or pseudoword) as within-subject factors. A main effect of language was found, $F(21, 106) = 8.16, p = .005, \eta^2 = .071$. As in previous studies, Portuguese-speaking children used a higher proportion of vowels. There was also a significant main effect of consonant–vowel structure, $F(1.8, 190) = 3.59, p = .034, \eta^2 = .033$, using the Greenhouse–Geisser correction for lack of sphericity, revealing that CVCV stimuli elicited a higher percentage of vowels, and there was an interaction between consonant–vowel structure and group, $F(2, 212) = 5.66, p = .004, \eta^2 = .051$, revealing that only the phonological group used more vowels for CVCV stimuli than for single-vowel stimuli. To rule out the possibility that the effects are due to the use of letters from children's own names, we performed a similar ANOVA entering percentage of vowels in children's names as a covariate. The main effect of language and the interaction between consonant–vowel structure and group remained significant. The finding that the prephonological group did not differ in its use of vowels across stimuli with different consonant–vowel structure confirms that these children are not influenced by phonology. The most important result is that the Portuguese prephonological spellers used a higher proportion of vowel letters than the U.S. prephonological spellers.

3.3.3. Bigram frequency

A child exposed to a particular language will see not only some letters more often than others but also some juxtapositions of letters more often than others. For example, *ee* is much more common in English words than in Portuguese words, and the opposite is true for *nh*. Are these differential frequencies reflected in the productions of even prephonological spellers? To address this issue, we counted the bigrams, or juxtapositions of two letters, in each language for both the book corpora and the lists of names. We found marked correlations between the lists, $\rho = .744$ for Portuguese and $\rho = .736$ for English, $p < .001$ for both. We then looked at how frequently the prephonological children used each bigram. Two separate by-item hierarchical regressions were carried out for each language, with the bigram being the unit of analysis and the dependent variable being the proportion of the time that children used the bigram. We made the distributions more normal by applying logarithmic transformations to the letter frequencies and the bigram frequencies in the corpora. In the first step, we entered the frequency of the two individual letters of the bigram. In the second step, the frequency of the bigram in the corpora was added. Our main question was whether the bigram frequency of the corpora would add significantly to the regression over and above the effect of the individual letter frequencies. In both languages, as Table 4 shows, bigram frequency accounted for a statistically significant amount of variance after frequencies of the individual letters had been entered. We found similar results when we ran the regression analyses looking only at those prephonological children whose names did not have the bigram in question.

Table 3
Percentages of vowels in children's spellings in English and Portuguese.

Consonant–vowel structure	English		Portuguese	
	Prephonological	Phonological	Prephonological	Phonological
CVC	.38 (.14)	.37 (.19)	.46 (.18)	.42 (.08)
CVCV	.37 (.12)	.40 (.16)	.45 (.21)	.50 (.07)
CCV	.37 (.14)	.37 (.17)	.46 (.18)	.43 (.10)

Note. Standard deviations are in parentheses.

Table 4

Summary of hierarchical regression analysis for variables predicting use of bigrams by prephonological spellers in English and Portuguese.

Variable	English				Portuguese			
	B	SE B	β	<i>p</i>	B	SE B	β	<i>p</i>
<i>Step 1</i>								
First letter frequency	.163	.024	.256	.000	.116	.012	.291	.000
Second letter frequency	.156	.024	.245	.000	.110	.012	.276	.000
<i>Step 2</i>								
Bigram frequency	.071	.012	.259	.000	.179	.013	.426	.000

Note. For English, $R^2 = .272$ for Step 1, $\Delta R^2 = .038$ for Step 2 ($p < .001$ for both). For Portuguese, $R^2 = .401$ for Step 1, $\Delta R^2 = .127$ for Step 2 ($p < .001$ for both). SE, standard error.

3.3.4. Consonant–vowel alternation

Another way to look at whether children are influenced by co-occurring letters is to examine alternation between consonants and vowels. As described earlier, even though the vowel–consonant distinction might not be present in a prephonological child, it aggregates information about a large number of letters in a way that the reader may find salient. Thus, we investigated whether the two languages differ in their alternation patterns and whether any such differences emerge in prephonological children's spellings. We asked first whether there was a higher proportion of consonant–vowel alternation in Portuguese text than in English. We have already shown that the ratio of consonants to vowels is much more equal in Portuguese than in English, and that fact alone would trivially predict more consonant–vowel alternation. To capture whether there is an additional trend toward alternation that cannot be accounted for by that difference in ratios, we conducted our analyses only between comparanda with the same proportion of vowels in them. This resulted in 22 different pairs of word groups, where each of the English words in a group had the same proportion of vowels as each of the Portuguese words in the same group. The words could have different consonant–vowel structures and lengths. For example, there was a group of words composed of 50% vowels that included *me* and *flea* in English and *bota* 'boot' and *umbigo* 'belly button' in Portuguese. Alternation was measured by looking at each adjacent pair of letters in the word and computing the average proportion of the time one of the letters was a vowel and the other letter was a consonant. The computations yield average consonant–vowel alternations of .70 for Portuguese and .65 for English. A *t* test performed across the groups showed a significant difference, $t(21) = 4.80$, $p < .001$. That is, even when controlling for the relative number of consonants and vowels in the words, Portuguese has more consonant–vowel alternation than English.

We performed the same analysis on the children's spellings. For the prephonological group, Portuguese-speaking children had a mean alternation proportion of .51 and English-speaking children had a mean alternation proportion of .48. For phonological spellers, the proportions were .62 and .58 for Portuguese and English speakers, respectively. A by-items ANOVA with the items being the groups of words that have the same proportion of vowels, language being the within-item factor, and speller group being the between-item factor showed only significant effect of language, $F(1,49) = 6.84$, $p = .012$, $\eta^2 = .123$; Portuguese spellers showed more alternation than English spellers. The effect of speller group did not reach significance ($p = .095$). These results suggest that children, even those who do not yet spell phonologically, are influenced by the patterns found in the textual environment.

3.4. Influences beyond the corpora

3.4.1. Children's own names

To investigate whether the children were influenced by the letters of their own names, we conducted analyses taking each letter of the alphabet as an item. For each letter, we counted the average number of times the letter was used by children with that letter in their forenames and compared that with the average number of times the same letter was used by children who did not have that letter in their forenames. Letters were dropped if no children had that letter in their names. This left 20 letters

Table 5

Average counts across letters for children who had the specific letter in their names and for children who did not have the letter in their names.

	English		Portuguese	
	Prephonological	Phonological	Prephonological	Phonological
With letter in name	13.50	11.63	10.83	8.57
Without letter in name	5.47	4.73	4.15	5.57

for English and 21 letters for Portuguese. Table 5 shows the average counts across letters for children who had the specific letter in their names and for children who did not have the letter in their names. Separate by-item ANOVAs were calculated for each language with presence of letter in the name and speller group as within-item variables. For both languages, presence of the letter in the child's name was a significant factor, with children being more likely to use the letters in their own names: for English, $F(1, 19) = 25.61$, $p < .001$, $\eta^2 = .574$; for Portuguese, $F(1, 20) = 50.08$, $p < .001$, $\eta^2 = .715$. For Portuguese, there was also a significant interaction, $F(1, 20) = 8.05$, $p = .01$, $\eta^2 = .287$, between presence of letter in the name and speller group, such that prephonological children were more likely to spell with letters in their own names.

The analyses just reported were carried out separately for English and Portuguese because the pool of letters that could be included was somewhat different in the two languages. To directly compare the results for the two languages, we discarded letters that were not found in participants' names in both countries, leaving 16 letters. We analyzed the ratio of the average letter use by children with that letter in their names to children without that letter in their names. For prephonological spellers, the average ratios across letters were 3.22 and 3.12 for English and Portuguese speakers, respectively. For phonological spellers, the ratios were 2.72 and 1.57 for English and Portuguese, respectively. An ANOVA on the ratios with language and speller group as within-item factors showed a significant effect of group, $F(1, 15) = 5.38$, $p = .035$, $\eta^2 = .264$. Prephonological children were more likely than phonological children to use letters of their names in their spellings. There was also a significant interaction between language and group, $F(1, 15) = 5.21$, $p = .037$, $\eta^2 = .258$, confirming that the difference between prephonological and phonological children was greater in the Portuguese-speaking group than in the English-speaking group.

3.4.2. Sequences from the alphabet

We investigated whether children display a tendency in their early productions to use letters in alphabetical order. Because some of the phonologically plausible spellings contain letter sequences in alphabetical order, we looked only at prephonological children in this analysis. A first step was to compute the number of bigram sequences, such as *ab*, *bc*, and *cd*, that were found in participants' spellings. We also needed to find the probability that children would use those bigram sequences by chance. To find this chance level, the letters in children's spellings were rearranged 10,000 times and the mean of the rearranged scores was computed. English-speaking nonphonological spellers had 7% of all their bigrams in alphabetic sequence, compared with an average 4% expected by chance. Portuguese-speaking nonphonological spellers had 5% of all their bigrams in alphabetic sequence, compared with an expectation of 3% by chance. A by-subject ANOVA with type of score (children's or rearranged) as a within-subject factor and language as a between-subject factor showed a significant effect only of score type, $F(1, 56) = 10.77$, $p = .002$, $\eta^2 = .161$. This result indicates that children were more likely to use alphabetic sequences than was expected by chance.

4. Discussion

The goal of this study was to investigate the nature of children's early spellings, which have been given short shrift in previous research. For example, why would a 4-year-old American write *sshidca* for *welcome home* (Bissex, 1980)? Phonologically oriented theories suggest that the earliest recognizable structures in young children's spellings come with the learning and use of letter-sound

associations; the child who wrote *sshidca* was simply stringing together some letters he knew in a random manner (Gentry, 1982). In contrast, constructivists (e.g., Ferreiro & Teberosky, 1982) propose that patterns appear even before children can match letters with sounds; for example, the child did not write *sssssss* because he knew that writing is not characterized by letters that repeat multiple times. We tested yet a third hypothesis, which is that children's early spellings reflect the knowledge they have gained from exposure to print in their environments. We call this view the statistical learning perspective because of its emphasis on learners' sensitivity to the statistical patterns in printed words.

Previous studies have provided tantalizing evidence that there are patterns even in spellings that predate the understanding that letters encode sounds systematically. For example, Bloodgood (1999) suggested that children who do not use letters to represent sounds overuse letters from their names. However, some aspects of those studies have not been conclusive. We believe that we have corrected many of these shortcomings. In previous studies, classification of children as phonological or not was based on experimenter judgment; there was no proof that the children under consideration were definitively prephonological. Here we have deployed a new test that objectively determines whether an individual child is a prephonological speller. We also developed more convincing methods for addressing the issue of the source of patterns—whether they are universalistic constructions or the result of statistical learning of environmental text. This was done by performing the same measurements on textual corpora—children's books and names—that we performed on children's spelling productions so as to see the extent to which they corresponded. Using two different languages that have the same script but different corpus statistics also served this goal because the statistical learning theory predicts differential spelling behavior occasioned by the differential textual environments. Finally, our randomization tests are a rigorous method of deciding whether observed patterns in child spellings are significant given other properties of the productions. For example, children who use only a few letters from the alphabet would be more likely to repeat letters or even whole "words" than would children who exploit the alphabet more fully; comparing their repetitions against what would happen if their own letters were randomized provides more compelling evidence of whether children intentionally avoid or embrace repetition.

Our primary question was whether the productions of the prephonological spellers reflected the nature of the text in their environments. We can sum up the evidence in the following paragraphs.

4.1. Spelling length

Even though three- and four-letter spellings were the most common types of production, children did not avoid one- and two-letter spellings nearly as often as the minimum quantity hypothesis would suggest. In fact, the proportion of one- and two-letter words in children's spellings was very similar to what is found in running texts. Brazilian children used short words—those with fewer than three letters—more often than U.S. children. This difference mirrors differences in the languages, with Portuguese having more short words than English.

4.2. Within-word variation

Evidence for statistical learning comes from the fact that prephonological spellers exposed to English text used geminates at approximately three times the rate of Portuguese spellers. This corresponds closely to the relative frequency of geminates in English (4–5% depending on corpus type) and in Portuguese (1–2%). When we compared children's productions against the rates that would be expected by chance levels (which are fairly high given that early writers tend to use a small set of letters over and over), we found a major disparity between languages; the difference between children's use of geminates and chance levels was significantly higher for Portuguese spellers than for English spellers, reflecting the fact that geminates are much more uncommon in Portuguese than in English.

4.3. Frequency of letters

The frequency distribution of letters in children's spellings correlated reasonably well with the distribution of letters in texts in their own environment. Also, Brazilian children used more vowel letters

than U.S. children at approximately the same proportion that Portuguese texts use more vowels than English texts.

4.4. Bigram frequency

The frequency distribution of bigrams in children's textual environment contributes significantly to predicting how often they write bigrams.

4.5. Consonant–vowel alternation

Brazilian children were somewhat more likely than U.S. children to alternate consonants with vowels in their spellings even after we controlled for the ratio of vowels to consonants. This mirrors the fact that Portuguese has more alternations than English.

4.6. Children's own names

Because children often see their own names written down, names constitute a particularly salient part of a child's textual environment. Prephonological spellers tend to use letters from their own names more often than children who do not have those particular letters in their names. This has been suggested before (e.g., Bloodgood, 1999), but here we showed that the results hold under a rigorous test of the prephonological status of the spellers in question.

4.7. Sequence of the alphabet

A special subcorpus that might be especially salient to beginning spellers is displays of the whole alphabet, which are almost always in alphabetical order. The prephonological spellers wrote contiguous letters of the alphabet in order nearly twice as often as would be expected if they wrote the same letters of their texts in random order.

Thus, many tests converge to indicate that provably prephonological spellers do not write letters at random, as one might surmise from considering their productions with the naked eye. Many formal aspects of these children's writing are influenced by formal characteristics of the textual environments in which they live. The characteristics we examined are very unlikely to be taught explicitly to preschoolers. That children pick up on these patterns supports the idea that implicit learning of patterns in the environment plays an important role even for young children who have not yet learned the basic principles of how writing works.

In addition to asking whether children's productions reflect the texts in their environment, we can ask whether the constructivists have been right in their claims about patterns in early spellings. We can sum up the evidence in the following paragraphs.

4.8. Minimum quantity

The absolute rates of short spellings—22% of all the Brazilians' productions—indicate that minimum quantity cannot be a strict rule. The observed differences between U.S. and Brazilian children in the rate of very short spellings and the strong correlation between the length distribution in children's spellings and text show that statistical properties of children's textual environment are important.

4.9. Within-word variation

Children use geminates less often than we would expect by chance, in line with the constructivist position, but they do not avoid them altogether. Indeed, the U.S. children wrote geminates in 13% of all the chances they had for putting two letters together. There was also a strong correlation with the text statistics of the children's linguistic environment, such that children exposed to English used geminates more often than children exposed to Portuguese. Again, an all-or-none principle posited by constructivists is only a tendency whose likelihood of application is correlated with language statistics.

4.10. *Between-word variation*

Constructivist theory teaches that children are careful to spell different words differently. We found, to the contrary, that children made more repetitions than was expected by chance even when their own limited selection of letters is taken into account.

4.11. *Syllabic spelling*

The hypothesis that young spellers go through a stage of writing one symbol per syllable is one of the best known components in the constructivist theory of spelling acquisition (Ferreiro & Teberosky, 1982), but some studies have cast doubt on it (Cardoso-Martins et al., 2006; Pollo et al., 2005). We again found no evidence in our sample; children used no more letters when spelling two-syllable stimuli than when spelling one-syllable stimuli. Prior reports of syllabic spellings in languages such as Portuguese and Spanish may reflect the letter name properties of the language; for example, Portuguese words have on average one letter name per syllable (Pollo et al., 2005). Syllabic spellings may be an incidental result of an attempt to use a letter name strategy. One may object that for our stimuli the desire to spell syllabically is opposed by a minimality requirement, such that children round up all spellings to three letters. However, as mentioned earlier, a good many of the observed spellings did consist of one or two letters.

In our view, a general statistical learning ability provides a parsimonious account not only for the scaled application of those constructivist patterns that we did find in our data but also for the additional patterns that we found such as attention to letter and bigram frequency. Consistent with the statistical learning view, children's spellings reflect patterns of text with which children have experience—books, their own names, and the alphabet listing.

Our results speak against the idea that early prephonological spellings are random and unpatterned. We saw correspondences between the spellings and certain characteristics of the print to which children are exposed—words, their own names, and the alphabet listing. Children's knowledge of some of these things helped to predict the accuracy of their spellings a year later. Although we found patterns in the early spellings, these were not always the patterns predicted by constructivists. When we did see a suggested pattern, such as a reluctance to repeat letters, there were serious doubts about its proposed universal aspect. Indeed, there were many differences between the spellings of the Brazilian and U.S. children. Thus, our findings suggest that the nature of early spellings is neither random nor universal.

All of the patterns that we found in the children's spellings can be explained by the same underlying argument: Children pick up recurring patterns of text even without knowing what those patterns mean in terms of letter–sound correspondence. The idea that children learn about subtle statistical regularities in the environment is not new (e.g., Perruchet & Pacton, 2006). For example, people use statistical regularities between co-occurring sounds in speech to help segment it into words (e.g., Saffran, Aslin, & Newport, 1996). A few studies have found evidence of statistical learning of formal patterns in spelling as well (Deacon et al., 2008). For example, children of approximately 6 years of age appear to know that certain letter sequences occur only in specific positions within an English word (Cassar & Treiman, 1997; Treiman, 1993) or that only certain letters may double in French (Pacton, Perruchet, Fayol, & Cleeremans, 2001). The children in these studies were probably using some phonological information in their spellings. The current study is the first, to our knowledge, to demonstrate that children can use important print information in their strictly prephonological spellings. Given a rigorous method of defining prephonological spellers, future studies may go on to examine other aspects of their spellings and how they predict later performance. Previous studies have not looked for individual differences in spellings at this level, assuming that all are random, but some spellings may be more patterned than others. This, in turn, should have practical implications for the early detection and remediation of children with literacy difficulties.

Acknowledgments

This work was supported in part by Grant R01 HD051610 from the National Institutes of Health. We are grateful to the teachers, parents, and children from the schools that participated in the study:

Colégio Marista Dom Silvério, Forsyth School, Washington University Nursery School, and University City Children's Center. This study is part of the PhD dissertation of the first author. Some of the findings were presented at the 15th annual meeting of the Society for the Scientific Study of Reading, Asheville, North Carolina, 2008.

Appendix A. Stimuli

A.1. Portuguese

CVC words: *bis* /bis/ 'encore', *diz* /dis/ 'say', *luz* /lus/ 'light', *mês* /mes/ 'month', *nós* /nɔs/ 'we', *voz* /vɔs/ 'voice'

CVCV words: *bule* /'buli/ 'teapot', *coma* /'kɔma/ 'eat', *dona* /'dɔna/ 'lady', *gado* /'gadu/ 'cattle', *gota* /'gota/ 'drop', *tela* /'tɛla/ 'screen'

CCV words: *crê* /kre/ 'believe', *cru* /kru/ 'raw', *pra* /pra/ 'to', *pré* /prɛ/ 'preschool', *pró* /prɔ/ 'in favor', *tri* /tri/ 'three times'

CVC pseudowords: /dɔs/, /has/, /lɔs/, /mɛs/, /nes/, /nis/

CVCV pseudowords: /'gɔba/, /'gɔfi/, /'hibu/, /'hɔmi/, /'nibu/, /'sofi/

CCV pseudowords: /bri/, /drɛ/, /kli/, /kra/, /ple/, /plo/

A.2. English

CVC words: *bite* /bait/, *dice* /daɪs/, *light* /laɪt/, *miss* /mɪs/, *nose* /noʒ/, *vote* /vot/

CVCV words: *pony* /'poni/, *coffee* /'kafi/, *daily* /'deɪli/, *goalie* /'goli/, *gummy* /'gʌmi/, *tummy* /'təmi/

CCV words: *blow* /blo/, *crow* /kro/, *draw* /dɔ/, *pray* /praɪ/, *tree* /tri/, *try* /traɪ/

CVC pseudowords: /kos/, /los/, /mis/, /nes/, /nis/, /ros/

CVCV pseudowords: /'gɔfi/, /'gomi/, /'nebo/, /'raɪbo/, /'ɔmi/, /'sɪfi/

CCV pseudowords: /braɪ/, /dɪi/, /klaɪ/, /kraɪ/, /plɑ/, /pɪu/

References

- Bissex, G. L. (1980). *Gnys at wrk: A child learns to write and read*. Cambridge, MA: Harvard University Press.
- Bloodgood, J. W. (1999). What's in a name? Children's name writing and literacy acquisition. *Reading Research Quarterly*, 34, 342–367.
- Cardoso-Martins, C., Corrêa, M. F., Lemos, L. 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.
- 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.
- Deacon, S. H., Conrad, N., & Pacton, S. (2008). A statistical learning perspective on children's learning about graphotactic and morphological regularities in spelling. *Canadian Psychology on Literacy Development*, 49, 118–124.
- Ehri, L. C. (1991). The development of reading and spelling in children: An overview. In M. Snowling & M. Thomson (Eds.), *Dyslexia: Integrating theory and practice* (pp. 63–94). London: British Dyslexia Association.
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9, 167–188.
- Evans, M. A., & Saint-Aubin, J. (2005). What children are looking at during shared storybook reading: Evidence from eye movement monitoring. *Psychological Science*, 16, 913–920.
- Ferreiro, E. (1990). Literacy development: Psychogenesis. In Y. Goodman (Ed.), *How children construct literacy: Piagetian perspectives* (pp. 12–25). Newark, DE: International Reading Association.
- Ferreiro, E., Pontecorvo, C., & Zuccheromaglio, C. (1996). Pizza or piza? How children interpret the doubling of letters in writing. In C. Pontecorvo, M. Orsolini, B. Burge, & L. Resnick (Eds.), *Children's early text construction* (pp. 145–163). Mahwah, NJ: Lawrence Erlbaum.
- Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. New York: Heinemann.
- Gentry, J. R. (1982). An analysis of developmental spelling in Gnys at wrk. *The Reading Teacher*, 36, 192–200.
- Gough, P. B., & Hillinger, M. L. (1980). Learning to read: An unnatural act. *Bulletin of the Orton Society*, 30, 179–196.
- International Phonetic Association (1999). *Handbook of the international phonetic association: A guide to the use of the International Phonetic Alphabet*. Cambridge, England: Cambridge University Press.
- Kamii, C., Long, R., Manning, M., & Manning, G. (1990). Spelling in kindergarten: A constructivist analysis comparing Spanish-speaking and English-speaking children. *Journal of Research in Childhood Education*, 4, 91–97.
- Kruskal, J. (1983). An overview of sequence comparison. In D. Sankoff & J. Kruskal (Eds.), *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison* (pp. 1–44). Reading, MA: Addison-Wesley.

- Levin, I., & Ehri, L. C. (2009). Young children's ability to read and spell their own and classmates' names: The role of letter knowledge. *Scientific Studies of Reading, 13*, 249–273.
- Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201–212.
- Martins, M. A., & Silva, C. (2001). Letter names, phonological awareness, and the phonetization of writing. *European Journal of Psychology of Education, 16*, 605–617.
- Nunes Carraher, T., & Rego, L. R. B. (1984). Desenvolvimento cognitivo e alfabetização [Cognitive development and the acquisition of literacy]. *Revista Brasileira de Estudos Pedagógicos, 63*, 38–55.
- 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.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10*, 233–238.
- Pinheiro, A. M. V. (1996). *Contagem de frequência de ocorrência de palavras expostas a crianças na faixa pré-escolar e séries iniciais do 1º grau* [Frequency counts of words presented to children in preschool and the initial series of first grade]. São Paulo, Brazil: Associação Brasileira de Dislexia.
- Pollo, T. C., Kessler, B., & Treiman, R. (2005). Vowels, syllables, and letter names: Differences between young children's spelling in English and Portuguese. *Journal of Experimental Child Psychology, 92*, 161–181.
- Pollo, T. C., Treiman, R., & Kessler, B. (2008). Preschoolers use partial letter names to select spellings: Evidence from Portuguese. *Applied Psycholinguistics, 29*, 195–212.
- Rego, L. R. B. (1999). Phonological awareness, syntactic awareness, and learning to read and spell in Brazilian Portuguese. In M. Harris & G. Hatano (Eds.), *Learning to read and write: A cross-linguistic perspective* (pp. 71–88). Cambridge, UK: Cambridge University Press.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
- Share, D. L., & Gur, T. (1999). How reading begins: A study of preschoolers' print identification strategies. *Cognition and Instruction, 17*, 177–213.
- Silva, C., & Alves-Martins, M. (2002). Phonological skills and writing of pre-syllabic children. *Reading Research Quarterly, 37*, 466–483.
- Sulzby, E. (1985). Kindergartners as writers and readers. In M. Farr (Ed.), *Advances in writing research, Volume 1: Children's early writing development* (pp. 127–199). Norwood, NJ: Ablex.
- Thompson, B. G., Cottrell, D. S., & Fletcher-Flinn, C. M. (1996). Sublexical orthographic relations early in the acquisition of reading: The knowledge sources account. *Journal of Experimental Child Psychology, 62*, 190–222.
- Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. New York: Oxford University Press.
- Treiman, R., Kessler, B., & Bourassa, D. (2001). Children's own names influence their spelling. *Applied Psycholinguistics, 22*, 555–570.
- Treiman, R., Kessler, B., & Evans, R. (2007). Anticipatory conditioning of spelling-to-sound translation. *Journal of Memory and Language, 56*, 229–245.
- Treiman, R., Kessler, B., & Pollo, T. C. (2006). Learning about the letter name subset of the vocabulary: Evidence from U.S. and Brazilian preschoolers. *Applied Psycholinguistics, 27*, 211–227.
- Treiman, R., & Rodriguez, K. (1999). Young children use letter names in learning to read words. *Psychological Science, 10*, 334–338.
- Vernon, S. A., & Ferreiro, E. (1999). Writing development: A neglected variable in the consideration of phonological awareness. *Harvard Educational Review, 69*, 395–414.
- Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.