

# Learning about the letter name subset of the vocabulary: Evidence from US and Brazilian preschoolers

REBECCA TREIMAN, BRETT KESSLER, and TATIANA CURY POLLO  
*Washington University, St. Louis*

Received: January 25, 2005      Accepted for Publication: June 13, 2005

## ADDRESS FOR CORRESPONDENCE

Rebecca Treiman, Campus Box 1125, Washington University, St. Louis, One Brookings Drive, St. Louis, MO 63130. E-mail: rtreiman@wustl.edu

## ABSTRACT

To examine the factors that affect the learning of letter names, an important foundation for literacy, we asked 318 US and 369 Brazilian preschoolers to identify each uppercase letter. Similarity of letter shape was the major determinant of confusion errors in both countries, and children were especially likely to interchange letters that were similar in shape as well as name. Errors were also affected by letter frequency, both general frequency and occurrence of letters in children's own names. Differences in letter names and letter frequencies between English and Portuguese led to certain differences in the patterns of performance for children in the two countries. Other differences appeared to reflect US children's greater familiarity with the conventional order of the alphabet. Boys were overrepresented at the low end of the continuum of letter name knowledge, suggesting that some boys begin formal reading instruction lacking important foundational skills.

A child's ability to identify the letters of the alphabet by name is one of the best predictors of how readily he or she will learn to read. Kindergarten letter identification accounts for nearly one-third of the variance in reading ability in Grades 1 to 3, and it is almost as successful at predicting later reading skill as an entire reading readiness test (Snow, Burns, & Griffin, 1998). Knowledge of letter names aids would-be readers and spellers in several ways (see Foulin, 2005). It helps them make some sense of printed words such as *jail*, where the entire name of one or more of the letters is heard in the spoken word. In addition, letter name knowledge helps children learn about the sound-symbolizing function of letters, because the phoneme that a letter represents is usually heard in the letter's name. Effects of letter name knowledge on reading, spelling, and letter sound knowledge have been documented in languages as diverse as English (McBride-Chang, 1999; Treiman, Tincoff, Rodriguez, Mouzaki, & Francis, 1998), Portuguese (Abreu & Cardoso-Martins, 1998), and Hebrew (Levin, Patel, Margalit, & Barad, 2002).

Given the foundational role of alphabet knowledge in literacy development, it is important to understand the processes involved in letter name learning. In the present study, we explore the hypothesis that these are the same processes

underlying the acquisition of spoken vocabulary in general. This hypothesis is motivated by the fact that children in the United States and other literate societies begin to learn the names of letters as early as 2 or 3 years of age, at the same time they are learning the names of many other things. For a young child, learning to label the shape *D* with the syllable /di/ may be quite similar to learning to label the shape  $\star$  with the label /star/.<sup>1</sup> It may be several years before the child realizes that *D* symbolizes a linguistic unit, a phoneme, and in this respect is different from  $\star$ . Letter name learning may thus form a bridge between the acquisition of a spoken vocabulary and the acquisition of literacy. If this hypothesis is correct, it would suggest that vocabulary development and literacy development are linked to one another in a way that has not previously been envisioned in the research literature.

To examine children's learning of letter names, and to determine whether it is affected by the same variables that influence vocabulary learning, we asked preschoolers in the United States and Brazil to identify each letter of the alphabet by name. We examined their correct responses and the nature of their errors. Uppercase letters were used because these are typically learned before lowercase letters (Worden & Boettcher, 1990). We tested a large number of children (318 in the United States, 369 in Brazil) to ensure sufficient power to detect effects. The cross-language comparison is useful because, although both English and Portuguese use the letters of the Latin alphabet, the languages differ in some important ways that were expected to be relevant to letter name learning. Table 1 shows one difference: the names that are given to the letters. The US and Brazilian names are similar in some cases but not others. For example, the name of *Q* rhymes with that of *B* in Portuguese but not in English. The languages also differ in the relative frequencies of the letters. For example, *T* is more common than *A* in English words, but the reverse is true in Portuguese. If children who speak different languages perform differently on the same letters, then the factors by which those letters differ in those languages are promising candidates for explaining children's performance. Researchers have documented some differences across languages and writing systems in the development of reading skills (e.g., Seymour, Aro, & Erskine, 2003), but we know of no direct cross-language comparisons of letter name learning.

Our hypothesis that children learn the names of letters in much the same way that they learn the nouns that label other concrete objects suggests that we should look to the literature on vocabulary development for suggestions about the factors that may be involved in letter name learning. This literature shows that shape plays an important role in word learning. Objects that are similar in shape often belong to the same category and are called the same name. Indeed, many of the count nouns in young children's vocabularies refer to classes of objects that are similar in shape (e.g., Gershkoff-Stowe & Smith, 2004). Children's reliance on shape is revealed in their generalizations about the names of objects. For example, one toddler said *moon* when playing with a half-moon shaped lemon slice and when touching a circular chrome dial on a dishwasher, having previously used this word when looking at the moon (Bowerman, 1978). Similar sorts of extensions occur with lowercase letters. For example, children may use the name of *b* for the similarly shaped *d* or the name of *p* for *q* (e.g., Courrieu & De Falco, 1989; Popp, 1964). Treiman and Kessler (2003), in regression analyses carried out over the 26 letters of the English alphabet, found that children made more naming errors on

Table 1. *Letter names in Brazilian Portuguese and American English*

Letter	Brazilian Portuguese	American English
A	/a/	/e/
B	/be/	/bi/
C	/se/	/si/
D	/de/	/di/
E	/ε/	/i/
F	/'εfi/	/εf/
G	/ʒe/	/dʒi/
H	/a'ga/	/etʃ/
I	/i/	/ai/
J	/'ʒota/	/dʒe/
K	/ka/	/ke/
L	/'εli/	/εl/
M	/'emi/	/εm/
N	/'eni/	/εn/
O	/ɔ/	/o/
P	/pe/	/pi/
Q	/ke/	/kju/
R	/'εhi/	/ar/
S	/'εsi/	/εs/
T	/te/	/ti/
U	/u/	/ju/
V	/ve/	/vi/
W	/'dabliw/	/'dʌblju/
X	/ʃis/	/εks/
Y	/'ipisilɔ/	/waɪ/
Z	/ze/	/zi/

lowercase letters that looked similar to many other letters than on lowercase letters that were not similar to many other letters. These effects were not significant with uppercase letters, however, raising questions about the role of shape for this set. In the current study, we analyze the entire letter name confusion matrix for uppercase letters rather than pooling the results for each letter and using the letter as the unit of analysis. We do not look at the overall error rate on *W*, for instance, but at how often children confuse *W* with particular letters such as *M* and *V*. With this more sensitive procedure, we expected that similarity of shape would be a major determinant of naming errors for uppercase letters, as for many of the objects in children's environments.

Just as the properties of a cat do not allow one to predict that its English name is *cat* as opposed to *potato*, one cannot predict the name of a letter from its shape. There are historical reasons why *B* is associated with the English name /bi/—the shape started out thousands of years ago as a stylized picture of a house, the word for which began with /b/ in a now extinct language—but these are not accessible to modern learners. Frequent repetition is required to fix arbitrary

associations in memory, and frequency of exposure plays a role in vocabulary learning generally (e.g., Schwartz & Terrell, 1983). We would expect it to be important in letter name learning as well. Surprisingly, Treiman and Kessler (2003) found that letter frequency in printed texts targeted at children did not contribute reliably to the prediction of preschoolers' correct responses on letter naming tasks. However, young children label the letters from their own first name more accurately than other letters (Levin & Aram, 2004; Treiman & Broderick, 1998; Treiman & Kessler, 2004), a finding that suggests that frequency may be largely a personal matter for young children. In the present study, we examined both general frequency effects—those that would be expected to hold across children—and individual frequency effects—those that reflect the occurrence of letters in specific children's names.

The present investigation of the effects of children's own names on their knowledge about letter names allows us to address an apparent discrepancy between the studies of Treiman and Broderick (1998) and Treiman and Kessler (2004), on the one hand, and the study of Levin and Aram (2004), on the other hand. In the former studies, which were carried out with English-speaking children, a benefit was found primarily for the first letter of a child's first name. In the latter study, which was carried out with Hebrew speakers, the advantage provided by own-name letters was not significantly larger for the first position of the name than for the second and third positions. Levin and Aram pointed to two factors that could cause Israeli children to show gains for more of the letters in their names. First, English capitalizes the first letter of proper names, which may make that letter particularly salient; Hebrew does not. This hypothesis cannot be addressed here because in Portuguese, like English, proper names are capitalized. Second, the spellings of Israeli names are relatively short. The shortness of the written names may encourage children to attend to all the letters rather than primarily the first letter. We used the Portuguese and English data to test Levin and Aram's hypothesis that the length of a child's name affects the pattern of own-name effects.

Another issue investigated in the present study concerns gender differences in letter name knowledge. Studies of vocabulary development have found that, during the early years, girls tend to have larger vocabularies than boys (e.g., Feldman et al., 2000). Girls show higher mean scores on reading tests later (e.g., Rutter et al., 2004). Some studies of literacy skills have also found boys to be more variable than girls, with gender differences most pronounced at the lower end of the distribution (e.g., Alexander & Martin, 2000). If the learning of letter names follows the same patterns as the learning of other words, gender differences may be found in early knowledge of letter names. The evidence on this point is mixed. Worden and Boettcher (1990) did not find statistically significant gender differences in letter naming and other tests of alphabet knowledge in a study with 180 US children between the ages of 2.5 and 7.5 years. Iversen, Silberberg, and Silberberg (1970), in their study of 110 US kindergartners, reported that girls could name more letters than boys. However, these researchers did not test the statistical significance of the differences. The present study, with many more participants than these previous studies, provides a new opportunity to ask whether girls and boys differ in their knowledge of letter names before formal reading instruction begins and, if so, whether they differ in central tendency, variability, or both.

Thus far, we have considered factors that would be expected to affect the learning of letter names and the learning of other vocabulary items in similar ways. Letters differ in some respects from other vocabulary items, though, and these differences may affect the learning process. One difference is that the names of letters are phonologically quite similar to one another in English and many other languages (see Treiman & Kessler, 2003). For example, English letter names rarely contain more than two phonemes, and they often end with /i/. The phonological similarities among the names of letters may help children identify these words as a class in a way that is not possible for such classes as animals or vehicles.

Although the letter names within a language are generally phonologically similar to one another, some pairs are especially close. For example, the English names of *P* and *B* are more similar than those of *P* and *Q*. Treiman and Kessler (2003) found that letters whose names share phonemes with many other letters led to more errors than letters whose names share phonemes with few other letters. In the present study, we examined phonological similarity effects at the level of stimulus–response (i.e., presented letter–erroneous answer) pairs rather than at the level of letters, testing the hypothesis that more similar pairs would be confused at higher rates. We also examined the interaction between phonological similarity and visual similarity, testing the hypothesis that letter pairs with similar shapes and similar names, such as *P* and *B*, would be especially prone to confusion. Researchers who have studied vocabulary development in general have noted confusions based on phonological similarity (e.g., Vihman, 1981), although these have not been a major focus of research. Such confusions may be especially prominent with letters because of their generally high similarity.

Letters differ from most other categories of words in that they have a conventional order. US children typically learn about this order from an early age. They can often sing or recite at least part of the alphabet in order by the age of 3, and performance is almost perfect by age 5 (Worden & Boettcher, 1990). Brazilian children, in our experience, have fewer opportunities to learn about the conventional order of the letters. Alphabet songs are less common in Brazil than the United States, and those that exist break up the ordered list of letters by providing a word that begins with each letter. For US children, at least, links may develop between pairs of adjacent letters, such as *F* and *G*. Children might sometimes interchange such letters' names, even when the letters are not especially similar to one another in other ways.

Several researchers have suggested that young children use information about objects' functions to name the objects when function is clearly linked to structure and when information about function is available at the time the categorization decision is made (e.g., Diesendruck, Hammer, & Catz, 2003; Kemler Nelson, Russell, Duke, & Jones, 2000). In the study by Kemler Nelson et al., for example, children who learned the label *tilfer* to refer to a gadget that rings a bell when a lever is pressed tended to extend that label to differently shaped gadgets that functioned similarly. However, preschool children are not usually taught the orthographic function of letters when they first learn to name them. Therefore, we did not expect function to play much role in young children's naming of letters.

Homophony is another variable that may play a somewhat different role in the learning of letter names than in the learning of other words. According to some

views of vocabulary development (e.g., Golinkoff, Mervis, & Hirsh-Pasek, 1994), the occurrence of a label that children have not previously heard cues them to connect that label with a previously unnamed object. When a sound sequence already has a meaning, children have difficulty learning a second meaning for that sequence (e.g., Backschneider, Gelman, Martinez, & Kowieski, 1999). Such difficulties may be less marked in the domain of letter names than in other domains, though, because the letter name meaning of a sequence such as /ti/ (*T*) is so different from its other meaning (*tea*).

To summarize, the present study was designed to examine the factors that affect children's learning of letter names and to compare them to the factors that are known to affect vocabulary learning in general. We examined these issues crosslinguistically by looking at letter naming by young children in the United States and Brazil.

## METHOD

### *Participants*

The US participants were 318 preschoolers from in and around Detroit, MI, all native speakers of English. At participating preschools we selected children who, based on their age, were scheduled to enter kindergarten in 1 or 2 years. The mean age of the US children was 4 years, 8 months (4;8, range = 3;8–5;9). The Brazilian participants were 369 preschoolers from the city of Belo Horizonte. We obtained data from children in the first year ( $N = 176$ ) and second year ( $N = 193$ ) of preschool. Their mean age was 5;1 (range = 3;3–6;9; ages were not available for three children). This age range is wider than in the United States because of variations in when children enter preschool and when they are promoted from one class to another. All of the US children and the large majority of the Brazilian children attended preschool programs that charged fees and that catered to middle-class families. In both countries, children from such backgrounds are exposed to letter names and letter shapes both at home and at preschool, typically in informal ways.

### *Procedure*

The children were shown each uppercase letter of the alphabet and were asked to say its name. Each letter was printed in large type on a separate card, and the cards were presented one at a time in a random order. Children were tested individually in a quiet location at their school, and all of the letters were presented in a single session. The experimenter, a native speaker of the child's language, encouraged the child to give a response for each letter. The experimenter provided general encouragement and praise, but did not indicate whether specific responses were correct or incorrect. The letters *K*, *W*, and *Y* were not presented to the Brazilian children because they are not part of the core spelling system of Portuguese, although they do appear in some foreign words and proper names. Diacritic marks, although common in Portuguese, were not used on any of the letters that were shown to the children.

Table 2. Mean proportion of responses of different types

Language	Correct Response	“Don’t know” or No Response	Incorrect Letter Name	Other Incorrect Response
English	.64	.18	.16	.02
Portuguese	.66	.19	.11	.04

## RESULTS

We began by breaking down the responses into four categories: correct response, “don’t know” or failure to respond, incorrect letter name, or some other type of error. Table 2 shows the proportion of responses in each country that fell into each of these categories. The majority of the “other” errors, in both countries, were numbers. Children sometimes labeled *I* as *one* or *B* as *eight*, for instance. These errors suggest a difficulty in distinguishing between the symbol systems of numbers and letters, a difficulty that is likely exacerbated by the visual similarities between the two systems. A few of the errors in the “other” category were descriptions of letters’ shapes, as when a Brazilian child called *I* *pauzinho* (little stick).

To compare the breakdown of the responses in the two countries, nonparametric tests were used. This was done because of indications of bimodality in the distribution of correct responses across children. Compared to the US children, the Brazilian children produced fewer errors that were real letter names ( $p = .032$  by a Mann–Whitney test, mean rank = 361 for US children and 330 for Brazilian children) and more other errors ( $p < .001$ , mean rank = 299 for US children and 383 for Brazilian children). These differences, although statistically significant, were small. In both countries, errors that were the names of letters greatly outnumbered errors that were not. The proportion of correct responses was slightly but not significantly lower for the US children than the Brazilian children, even though the US children were significantly younger than the Brazilian children,  $t(682) = 9.66$ ,  $p < .001$ ,  $SE$  difference = 0.50. This difference is consistent with our observation that more stress is placed on early learning of the alphabet in the United States than in Brazil. The fact that the two groups performed at similar levels gives us the opportunity to compare the factors that contribute to performance in the two countries for children who are equated for overall levels of correctness.

Given our interest in the factors that cause children to confuse letters with one another, we tabulated the number of errors to each letter in each language that involved each other letter as a response. We did not analyze errors involving *K*, *W*, or *Y* in Portuguese because, as stated earlier, these letters do not occur in the native words of Portuguese and were not presented to the Brazilian children. Regression analyses were carried out for each language to predict confusions on each stimulus–response pair from a variety of factors. For example, one of the most common errors among both US and Brazilian children was to respond with the name of *B* when shown *D*. Our analyses were designed to uncover the properties

of the letters and letter pairs that are associated with this and other confusions. Confusions on *B* and *D*, for example, may arise because these letters are similar in appearance and have similar-sounding names.

To assess the visual similarity of pairs of letters, we had 30 US college students rate the visual similarity of all pairs on a scale from 1 (*not at all similar*) to 7 (*very similar*). There were 10 different random orders, and three participants were assigned to each order. Across participants, half saw the letters of a pair in one order and half saw them in the other order. The letters were presented in a font that was similar to that seen by the children. The participants' visual similarity ratings correlated highly ( $r = .83$ ) with those collected by Boles and Clifford (1989), although we had 30 participants rate each pair and they had only 12. Our ratings also correlated ( $r = .71$ ) with number of shared features according to the scheme of Briggs and Hocevar (1975), which is based on an analysis of the visual features of uppercase letters that hold across a variety of fonts.

As a measure of phonological similarity, we counted the number of phonemes that pairs of letter names in each language share in the same position. For example, the English name of *P*, /pi/, shares a phoneme with the name of *B* but does not share any phonemes with the name of *Q*. The Portuguese name of *P*, /pe/, shares a phoneme with the name of *B*, /be/, and a phoneme with the name of *Q*, /ke/. Because relatively few letter names share two phonemes, especially in English, we coded the pairs according to whether they shared zero phonemes versus one or more phonemes.

Other things being equal, children may make fewer confusion errors when presented with common letters than uncommon letters. In addition, children's erroneous responses may tend to be the names of common letters. To derive a measure of letter frequency in English that would be appropriate for young children, we examined the words that appear with a *U* value (frequency per million words adjusted for variation in distribution of words across content areas) of one or more in written materials at the kindergarten and first-grade levels in Zeno, Ivens, Millard, and Duvvuri (1995). The analyses were limited to words that also appeared in a second source (CMU Pronouncing Dictionary, 1998), eliminating some erratic or low-frequency words. We calculated how often each letter occurred in this corpus, weighting words by their frequencies, and coded the frequency of the stimulus and response letter in each pair. A similar approach was used for Portuguese, based on the words that appear in the preschool and first-grade corpus of Pinheiro (1996) and ignoring diacritics. Our assumption here is not that these preschoolers can read (the large majority cannot), but that letters that occur more frequently in written materials are likely to be discussed more often by parents and preschool teachers and are likely to be of more interest to children.

For *A*, *B*, and *C*, counts of text frequency may underestimate children's exposure to the letters. In both English and Portuguese, the alphabet is commonly called the *ABCs*, and children have extra experience with these letters' names and shapes. Children may be influenced by this extra exposure, making fewer confusions when shown *A*, *B*, or *C*, producing more errors that are the names of *A*, *B*, or *C*, or both. We therefore coded whether the stimulus letter and response letter of each pair belonged to the *ABC* set. To test whether children tended to confuse letters



Table 3. *Standardized regression coefficients for variables in regressions predicting confusions between pairs of letters*

Variable	English	Portuguese
Visual similarity of letters' shapes	.32***	.27***
Phonological similarity of letters' names	.02	.05
Visual Similarity × Phonological Similarity	.19*	.23*
Stimulus letter frequency	-.02	-.07
Response letter frequency	.07*	.18***
Stimulus letter A, B, or C	-.11***	-.06
Response letter A, B, or C	.10**	.19***
Adjacency of letters in alphabet	.12***	-.02

Note: Total  $R^2 = .23$  for English and  $.30$  for Portuguese ( $ps < .001$ ).  
\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

that are next to one another in the conventional alphabet sequence, we also coded whether the stimulus letter and the response letter in each pair are adjacent to one another in the alphabet.

Before running regressions, we examined the distributions of the variables. Log transformations were applied to the variables of number of confusions, similarity rating, and stimulus and response letter frequency to make the distributions more normal. Separate regression analyses were carried out for each language. The first stage of the regressions used the variables of visual similarity of letter shapes, phonological similarity of letter names, frequency of stimulus letter, frequency of response letter, *ABC* status of stimulus letter, *ABC* status of response letter, and adjacency in the alphabet. We then introduced the interaction between shape similarity and name similarity, calculated as the product of the two variables. This interaction added significantly to the regression for each language ( $p = .044$  for English,  $p = .021$  for Portuguese), and Table 3 shows the results when it was included.

In both languages, visual similarity of letter shapes was positively associated with confusions. Children tended to interchange letters that were similar in appearance, such as M and W. Indeed, similarity of letter shapes was the strongest single determinant of confusion errors in both English and Portuguese. Although phonological similarity of letter names had a significant effect in both languages when the interaction term was not included in the regressions, Table 3 shows that this effect was almost completely eliminated when the interaction between visual and phonological similarity was included. The interaction was second in strength only to the main effect of visual similarity. This outcome indicates that children were especially likely to interchange letters that had similar names as well as similar shapes. For example, B and D are similar in both dimensions in English and Portuguese, and children often confused them. C and T are similar in name but dissimilar in shape, and they were less often confused. Variables related to letter frequency had significant effects as well. Both US and Brazilian children were significantly more likely to produce a letter as a response when that letter was frequent in their language than when it was infrequent. Frequent stimulus

letters tended to give rise to less confusion than infrequent stimulus letters, but this trend was not significant for either group of children. When children made an error, they were more likely to respond with the names of *A*, *B*, or *C* than expected on the basis of other factors. This effect was reliable for both the US and Brazilian children. The tendency to produce fewer confusions when shown *A*, *B*, or *C* than when shown other letters was statistically reliable in English but not in Portuguese. Finally, the US children were significantly more likely to confuse letters that were adjacent to one another in the alphabet than letters that were not adjacent. For example, these children sometimes called *F* by the name of *G*, even though these letters are not very similar in their shapes or names. For the Brazilian children, adjacency in the alphabet did not have a significant effect.

To determine whether children have difficulty with letters such as *T*, whose labels are already known in another context (*tea*), we added to the regression analyses the frequency of the stimulus letter name as a word according to the corpora mentioned above. This variable did not add significantly to the regression for either language. For English, this held true whether or not *A* was counted as a word. (Although the full, stressed pronunciation of the word *a* is /e/, like the name of the letter, the word is not usually pronounced this way.)

The linguistic function of a letter is not related to its shape, and we did not expect function to play much role in letter naming. To test this idea, we looked at the most basic distinction in function, that between consonant and vowel. In neither language did agreement in consonant–vowel status account for significant additional variance. For English, this held true when we omitted *Y* from the analyses, because it can serve as either a consonant or vowel, and when we counted it as agreeing in status with both vowels and consonants. As a further test, we asked whether children produced errors that followed a common phonological pattern for letter names in their language and that were based on a common sound of the letter, for letters whose names do not contain that sound. For example, if an English-speaking child called *H* /hi/, this would suggest that the child constructed a name for the letter based on its sound-symbolizing function. We observed only one such error, that of a Brazilian child who called *G* /ga/. This error may, in part, reflect the child's knowledge that *G* can symbolize /g/, a phoneme that is not included in the letter's Portuguese name. Overall, though, the results suggest that the preschoolers rarely attended to letters' functions in the naming task.

For some pairs of letters, we expected English and Portuguese speakers to show different patterns of errors based on the different properties of the letters in the two languages. As mentioned previously, the phonological similarity of certain pairs of letter names differs in Portuguese and English, and the relative frequencies of letters differ across the two languages. One way to test the hypothesis that children are affected by the properties of the letters in the language to which they have been exposed is to carry out regressions for one language using the predictors appropriate for the other language. This is a weak test, as the phonological similarity of pairs of letter names correlates across English and Portuguese ( $r = .59$ ), as does letter frequency ( $r = .84$ ). Nevertheless, the proportion of variance explained by the regressions declined for both English and Portuguese when the predictors from the other language were used.

Table 4. Mean number of confusions involving pairs of letters for which the phonological similarity of names differs in English and Portuguese

Language Tested	Language in Which Letter Names Are Similar	
	English	Portuguese
English	2.50	2.31
Portuguese	0.96	1.92

Note: Letter names are considered similar in a language if they share at least one phoneme in the same position.

As a stronger test of the language specificity of letter name effects, we examined the results for those pairs of letters for which the phonological similarity of the names differs in English and Portuguese. These are letters that rank high on the interaction term involving phonological similarity and visual similarity in one language but not the other. As Table 4 shows, the US children tended to make more errors involving letters whose names share one or more phonemes in the same position in English but not in Portuguese. Conversely, the Brazilian children tended to make more errors involving letters whose names share phonemes in Portuguese but not English. This pattern was reflected in a significant interaction in an analysis of variance (ANOVA) using the factors of language and pair type,  $F(1, 74) = 5.27, p = .024, \eta_p^2 = .067$ . The ANOVA also showed a main effect of language,  $F(1, 74) = 15.14, p < .001, \eta_p^2 = .170$ . This main effect arose because the proportion of errors that were real letter names was somewhat higher among the US children than the Brazilian children, as noted earlier, even though the proportion of correct responses was not higher among the US children.

The trends observed in the regressions suggest that children make more errors in which they call a less common letter by the name of a more common letter than the reverse. To test the hypothesis that such errors are governed by the frequencies of the letters in the language to which the children have been exposed, we examined those letter pairs for which the relationship between the frequency of the stimulus letter and the frequency of the response letter differs in the two languages. One such pair, as mentioned previously, is *A* and *T*: *T* is more common than *A* in English, but *A* is more common than *T* in Portuguese. Table 5 shows the results for pairs such as these. An ANOVA using the factors of pair type and language found a significant interaction,  $F(1, 98) = 6.54, p = .012, \eta_p^2 = .063$ . The English speakers tended to make more errors where the response letter was more frequent than the stimulus letter in English, and the Portuguese speakers made more errors where the response was more frequent than the stimulus in Portuguese. That is, the results for the same letter pairs differed as a function of the letters' frequencies in the written materials of the language. The ANOVA also showed a main effect of language,  $F(1, 98) = 8.56, p = .004, \eta_p^2 = .081$ . This arose because the proportion

Table 5. Mean number of confusions involving pairs of letters for which the relationship between stimulus letter frequency and response letter frequency differs in English and Portuguese

Language Tested	Language in Which Response Letters Are More Frequent Than Stimuli	
	English	Portuguese
English	2.60	2.18
Portuguese	1.42	2.10

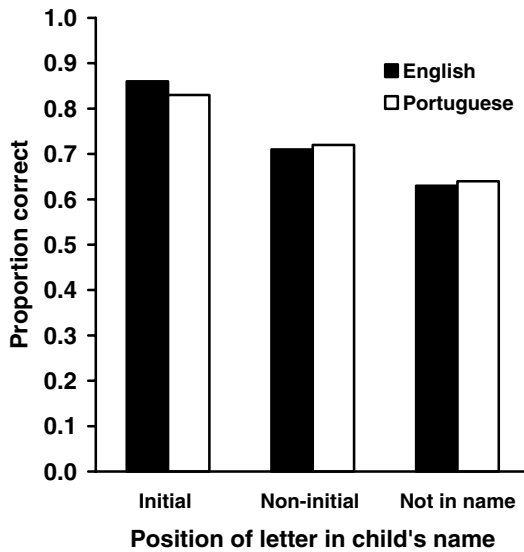


Figure 1. The mean proportion of correct responses on letters for children who have the letter in the initial position of the first name, children who have the letter in a noninitial position of the first name, and children who do not have the letter in the name.

of errors that were letter names was higher for the US children than the Brazilian children.

Thus far, we have examined effects of letter frequency that would be expected to hold across children in a particular society. To examine letter frequency effects that are specific to individual children, we tabulated the results for each letter for children who had that letter as the first letter of their first name or commonly used nickname, children who had that letter in a noninitial position of their name, and children who did not have the letter in their name. The results appear in Figure 1.

For both languages, the status of the letter in the child's name had a significant effect: English:  $F(2, 16) = 20.45$ ,  $\eta_p^2 = .719$ ; Portuguese:  $F(2, 26) = 16.35$ ;  $\eta_p^2 = .557$ ;  $p < .001$  for both; the analyses of own-name effects included only those letters for which the denominators for all of the proportions were at least nine to ensure that proportions could be calculated reliably. The superiority for initial over noninitial letters was statistically reliable for each language, as was the superiority for noninitial over not-in-name letters. When we confined our analyses to children who scored in the lowest quartile of proportion correct on the letter name task, the effects of the children's names were even larger than those shown in Figure 1. For this group, the proportion of correct responses was .55 for name-initial letters, .19 for later letters, and .10 for not-in-name letters.

Levin and Aram's (2004) hypothesis that children with short names tend to pay attention to all of the letters in their name predicts that the difference between letters later in the name and letters not in the name will be larger for children with short names than for children with long names. As one test of this hypothesis, we divided the total group into those with shorter names (six or fewer letters,  $N = 379$ ) and those with longer names (more than six letters,  $N = 308$ ). The difference in proportion correct between the later-in-name case and the not-in-name case was numerically larger for the children with short names than for the children with long names (.08 vs. .05). However, the interaction did not approach significance in an ANOVA using the factors of letter's status in name and name length, pooling the results over the two languages. As a further test, we omitted the results for children who had names of six or seven letters and looked at the extreme groups. The short name group in this analysis had names that, at five or fewer letters, were more similar to the lengths of Israeli children's names. Again, the critical interaction was not significant.

We carried out additional analyses to determine whether boys and girls differ in knowledge of letter names, as they do in general vocabulary knowledge. The mean proportion of correct responses was higher for girls than boys in each country, but not significantly so by a Monte Carlo test based on 10,000 rearrangements (English:  $p = .088$ , one tailed; Portuguese:  $p = .169$ , one tailed). Pooling over the two countries, the mean proportion of correct responses was .67 for girls and .63 for boys, a difference of about one letter. In each country, boys' scores were significantly more extreme than girls' ( $p < .001$  by a Moses test). More boys than girls fell at the low end of the distribution. In the combined results for the two countries, for example, 20% of the boys had a score of less than 20% correct on the letter name task, compared to 15% of the girls.

## DISCUSSION

In the work reported here, we explored how children learn the labels for one class of items, letters. The results support our hypothesis that the early learning of letter names follows many of the same principles that apply to vocabulary learning in general, the operation of these principles reflecting the characteristics of the letter name subset of the vocabulary.

Both letter names like /di/ and words like *cat* apply to a range of objects. An important part of vocabulary learning is learning which objects should be placed in the same category and given the same label. With many everyday objects, these categories are defined partly on the basis of shape. This is even more true for letters. Children's attention to shape in the case of letters is shown by the finding that they sometimes misidentified a letter as another letter that is similar to it in shape. This was so even though no uppercase letters are mirror images of one another, as with the lowercase pairs that have been reported to cause particular difficulties for children (e.g., Courrieu & De Falco, 1989; Popp, 1964). Some of the more common shape-based confusions in our study involved letters that are similar when inverted, such as M-W and A-V. Other confusable letter shapes differ in the closing of a curve (C to O) or the addition of a small line (O-Q). Pairs such as O-Q may be difficult, in part, because adding a short line sometimes does not change a letter's category (as when changing from a sans serif font to a serif font), but at other times does change the category.

The phonological characteristics of letter names affected children's performance as well. Children tended to confuse letters with similar names when they also had similar shapes. For example, children sometimes mislabeled *P* as *B* and vice versa. Moreover, error patterns on the same pairs of letters sometimes differed for Portuguese and English speakers, reflecting the different names of the letters in the two languages. Confusions on the basis of sound occur in the learning of other vocabulary words as well (e.g., Vihman, 1981). However, such confusions may be especially noticeable with letters because many letters in English, Portuguese, and other languages are similar in name as well as shape.

Both letter names like /di/ and words like *cat* cannot be predicted from the referent's shape. With arbitrary stimulus-response associations such as these, repetition is needed to secure the links in memory. Speaking to the important role of frequency in vocabulary learning, the earliest learned words tend to be ones that occur often in children's environments, and the child's own name is one of the first sound patterns to be recognized (Mandel, Jusczyk, & Pisoni, 1995). Frequency affected letter name learning as well, in that children's errors were significantly more likely to be letters that are common in their language than letters that are less common. Children also tended to make fewer confusion errors when shown frequent letters than when shown less frequent letters, but this trend was not statistically reliable. The effects that we found for *A*, *B*, and *C* can also be interpreted as frequency effects. Children tended to produce these letters as errors and, in the United States, they tended to make fewer confusions when these letters were presented to them. Children receive additional exposure to *A*, *B*, and *C* beyond their frequency in texts, as when these letters appear on alphabet teaching devices, and this additional exposure seems to affect their performance.

In addition to general frequency effects, we found personal frequency effects that reflected the letters in children's first names. Previous studies have found that English-speaking children are better at labeling the first letter of their name than letters that do not appear in their name (Treiman & Broderick, 1998; Treiman & Kessler, 2004). The same was true here, for Brazilian children as well as US children. We also saw a small but statistically significant advantage for the later letters in the child's name relative to not-in-name letters. Of particular importance

are the large effects of name membership that we saw for children in the lowest quartile of performance. The results suggest that children often begin by learning the name of the first letter of their own first name and perhaps a few other common letters. Children then learn the remaining letters of their name and the other letters of the alphabet rather quickly. This pattern can explain why a number of children in the present study knew the names of only a few letters, relatively few children knew an intermediate number of letters, and many children knew all or almost all.

Studies of general vocabulary learning have reported differences between boys and girls (e.g., Feldman et al., 2000). Similarly, we saw gender differences in children's knowledge of letter names. The mean scores for boys were slightly but not significantly lower than those for girls, and boys were significantly overrepresented at the low end of the distribution. Our results do not shed light on the reasons for the gender differences, but they indicate that such differences are not restricted to one subset of the vocabulary. Our results further suggest that, when formal literacy instruction begins, more boys lack the knowledge of letter names that can aid reading and spelling performance. This difference may help explain why more boys than girls score at the low end of the distribution in tests of reading ability in school (e.g., Alexander & Martin, 2000). Our results do not support the suggestion that boys' classroom behaviors and teachers' reactions to those behaviors suffice to explain the observed gender differences in reading ability (Prochnow, Tunmer, Chapman, & Greaney, 2001).

Although the learning of letter names appears to be similar in many ways to the learning of other vocabulary items, letter names have some unique characteristics. In most languages, including the ones studied here, the names of letters sound quite similar to one another. The phonological similarity of letter names may help children pick out the names as a set, facilitating learning. Letters also differ from the members of most other semantic categories in having a conventional order. The US children in our study, although not the Brazilian ones, showed a reliable tendency to confuse letters that were next to one another in the alphabet, even letters that were not especially similar to one another in other ways. US children learn the conventional order of the alphabet from an early age, via a song, and this may make letter ordering quite salient for them.

Certain variables that have been reported to affect the learning of other vocabulary items did not have noticeable effects on letters. One such variable was homophony (e.g., Backschneider et al., 1999). The children in our study did not have special problems when the name of a letter was a phonological string with another meaning. The contexts in which children learn about letters may be different enough from the contexts in which they learn about other sorts of words that any negative effects of homophony are minimal. Function, like homophony, may play a smaller role with letters than with other vocabulary words. As mentioned previously, young children who do not know the names of objects sometimes apply the names of other objects of similar function (e.g., Kemler Nelson et al., 2000). However, our data suggest that preschoolers do not do this with letters. Children may use function as a basis for the classification and naming of objects only when the objects' functions are clear to them. Preschool children, however, are not very knowledgeable about the sound-symbolizing function of letters.

Overall, our results show that the early learning of letter names and the early learning of other words are similar in many ways, and that letter name learning occurs in similar ways across languages. These similarities reflect the arbitrary nature of the labels and the rote memorization that is required to learn them. The payoff for this learning is different in the domain of letters and the domain of other objects, though. Learning the label *cat* (English) or *gato* (Portuguese) does not teach children anything about cats. In contrast, learning the label /di/ (English) or /de/ (Portuguese) for *D* helps children learn something new about *D*. Because the name of a letter generally contains the sound that the letter symbolizes, the fact that *D* stands for /d/ becomes motivated rather than arbitrary once the letter name is learned. Indeed, children appear to use letter names to help learn and remember letter sounds and begin to make sense of words' spellings (e.g., McBride-Chang, 1999; Treiman et al., 1998). In this way, letter learning forms a bridge between early vocabulary learning and the later acquisition of literacy.

### ACKNOWLEDGMENTS

This research was partially supported by NSF Grant BCS-0130763. We are grateful to Cláudia Cardoso-Martins for sharing data from Brazilian children. Thanks to Margo Bowman, Rochelle Evans, and Shannon Ross for their help in data collection and analysis.

### NOTE

1. The phonemes are represented using the alphabet of the International Phonetic Association (1996, 1999).

### REFERENCES

- Abreu, M. D., & Cardoso-Martins, C. (1998). Alphabetic access route in beginning reading acquisition in Portuguese: The role of letter-name knowledge. *Reading and Writing: An Interdisciplinary Journal*, *10*, 85–104.
- Alexander, J. R. M., & Martin, F. (2000). Norming tests of basic reading skills. *Australian Journal of Psychology*, *52*, 139–148.
- Backschneider, A. G., Gelman, S. A., Martinez, I., & Kowieski, J. (1999). Children's use of different information types when learning homophones and nonce words. *Cognitive Development*, *14*, 515–530.
- Boles, D. B., & Clifford, J. E. (1989). An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, Instruments, and Computers*, *21*, 579–586.
- Bowerman, M. (1978). The acquisition of word meaning: An investigation into some current conflicts. In N. Waterson & C. Snow (Eds.), *Developmental of communication: Social and pragmatic factors in language acquisition* (pp. 263–287). New York: Wiley.
- Briggs, R., & Hocevar, D. J. (1975). A new distinctive feature theory for upper case letters. *Journal of General Psychology*, *93*, 87–93.
- CMU Pronouncing Dictionary (Version 0.6). (1998). Retrieved July 19, 2004, from the Carnegie Mellon University, Speech at CMU, <ftp://ftp.cs.cmu.edu/project/speech/dict/cmudict.0.6> [Data file].
- Courrieu, P., & De Falco, S. (1989). Segmental vs. dynamic analysis of letter shape by preschool children. *Cahiers de Psychologie Cognitive*, *9*, 189–198.
- Diesendruck, G., Hammer, R., & Catz, O. (2003). Mapping the similarity space of children's and adults' artifact categories. *Cognitive Development*, *18*, 217–231.
- Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development*, *71*, 310–322.



- Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing, 18*, 129–155.
- Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development, 75*, 1–17.
- Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language, 21*, 125–155.
- International Phonetic Association. (1996). *Reproduction of The International Phonetic Alphabet*. Retrieved May 17, 2005, from <http://www.arts.gla.ac.uk/IPA/ipachart.html>.
- International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- Iversen, I. A., Silberberg, N. E., & Silberberg, M. C. (1970). Sex differences in knowledge of letter and number names in kindergarten. *Perceptual and Motor Skills, 31*, 79–85.
- Kemler Nelson, D. G., Russell, R., Duke, N., & Jones, K. (2000). Two-year-olds will name artifacts by their functions. *Child Development, 71*, 1271–1288.
- Levin, I., & Aram, D. (2004). Children's names contribute to early literacy: A linguistic and a social perspective. In D. Ravid & H. Bat-Zeev Shyldkrot (Eds.), *Perspectives on language and language development* (pp. 219–239). Dordrecht: Kluwer.
- Levin, I., Patel, S., Margalit, T., & Barad, N. (2002). Letter names: Effect on letter saying, spelling, and word recognition in Hebrew. *Applied Psycholinguistics, 23*, 269–300.
- Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science, 6*, 314–317.
- McBride-Chang, C. (1999). The ABC's of the ABC's: The development of letter-name and letter-sound knowledge. *Merrill-Palmer Quarterly, 45*, 285–308.
- 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 of occurrence counts of words exposed to children at the preschool level and the initial years of primary school]. São Paulo, Brazil: Associação Brasileira de Dislexia.
- Popp, H. M. (1964). Visual discrimination of alphabet letters. *The Reading Teacher, 17*, 221–226.
- Prochnow, J. E., Tunmer, W. E., Chapman, J. W., & Greaney, K. T. (2001). A longitudinal study of early literacy achievement and gender. *New Zealand Journal of Educational Studies, 36*, 221–236.
- Rutter, M., Caspi, A., Fergusson, D., Horwood, L. J., Goodman, R., Maughan, B., et al. (2004). Sex differences in developmental reading disability: New findings from 4 epidemiological studies. *Journal of the American Medical Association, 291*, 2007–2012.
- Schwartz, R. G., & Terrell, B. Y. (1983). The role of input frequency in lexical acquisition. *Journal of Child Language, 10*, 57–64.
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundations literacy acquisition in European languages. *British Journal of Psychology, 94*, 143–175.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Treiman, R., & Broderick, V. (1998). What's in a name: Children's knowledge about the letters in their own names. *Journal of Experimental Child Psychology, 70*, 97–116.
- Treiman, R., & Kessler, B. (2003). The role of letter names in the acquisition of literacy. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 31, pp. 105–135). San Diego, CA: Academic Press.
- Treiman, R., & Kessler, B. (2004). The case of case: Children's knowledge and use of upper- and lowercase letters. *Applied Psycholinguistics, 25*, 413–428.
- Treiman, R., Tincoff, R., Rodriguez, K., Mouzaki, A., & Francis, D. J. (1998). The foundations of literacy: Learning the sounds of letters. *Child Development, 69*, 1524–1540.
- Vihman, M. M. (1981). Phonology and the development of the lexicon: Evidence from children's errors. *Journal of Child Language, 8*, 239–264.
- Worden, P. E., & Boettcher, W. (1990). Young children's acquisition of alphabet knowledge. *Journal of Reading Behavior, 22*, 277–295.
- Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.