Research Article

Novel Word Learning in Children Who Are Bilingual: Comparison to Monolingual Peers

Mary Alt,^a Genesis Dominique Arizmendi,^a Shelley Gray,^b Tiffany Patrice Hogan,^c
Samuel Green^{b†}, and Nelson Cowan^d

Purpose: We compared novel word learning in 2nd-grade children with typical development who were Spanish–English bilinguals to English monolinguals to understand word learning in bilingual children.

Method: Children (monolinguals $n = 167$, bilinguals $n = 76$) engaged in 5 computer-based tasks that assessed word learning in 6 different contexts. The tasks measured children's ability to link novel names with novel objects/actions, make decisions about the accuracy of those names and objects/actions, recognize the semantic features of the objects/actions, and produce the novel names. For analysis, we used Bayesian repeated-measures analyses of covariance with Bayesian independent-samples t tests to clarify interactions.

Results: Monolingual and bilingual children differed in some, but not most, word learning situations. There was at least moderate evidence that bilingual children were less accurate at naming in 1 condition and at detecting mispronunciations in 3 of 6 contexts and were less accurate at judging semantic features of a referent when that referent was paired with orthographic information.

Discussion: Among children with typical development, there were few differences in novel word learning between monolingual and bilingual participants. When differences did occur, they suggested that bilinguals were more accepting of phonological variations of word productions than their monolingual peers.

The purpose of this study was to compare novel word learning in bilingual (Spanish–English) school-age children with typical development to age-matched monolingual English-speaking children with typical development. Our main research question was whether we would find word learning differences in typically developing monolingual and bilingual children on novel word learning tasks that used English phonology. We know very little about how school-age bilingual children learn new words. Indeed, Buac, Gross, and Kaushanskaya (2016) note that this literature is “sparse” (p. 24). One reason may be that we do not expect word learning deficits precisely because these children have typical development (e.g., Byers-Heinlein, Fennell, & Werker, 2013; De Houwer, Bornstein, & Putnick, 2014). However,

the literature makes clear that vocabulary growth is different for English and Spanish for bilingual children (e.g., Rojas & Iglesias, 2013) and is different from the growth of vocabulary for monolingual children. One major reason for the difference is that bilingual children are distributing their knowledge across two languages that do not have translational equivalents for every word (e.g., Hoff et al., 2012).

The bulk of the existing literature comes from infants, toddlers, and preschoolers, making it difficult to understand word learning in older school-age children. However, there have been many advances in the field to give us insight into bilingual word learning across the life span. For example, there is evidence that, in general, bilingual adults have an advantage in learning words from a foreign language (e.g., Kaushanskaya & Marian, 2009; Nair, Biedermann, & Nickels, 2016) but may know fewer vocabulary words than monolinguals in any one language (e.g., Bialystok & Luk, 2012). On the other end of the spectrum, we know that monolingual and bilingual infants (e.g., Houston-Price, Caloghris, & Raviglione, 2010) and toddlers (Brojde, Ahmed, & Colunga, 2012) approach word learning differently. Blumenfeld and Marian (2009) “consider childhood bilingualism as a special case of language acquisition” (p. 2). The literature on word learning in very young children and adults is important and valuable in understanding word learning across the life span. However, in our

^aThe University of Arizona, Tucson

^bArizona State University, Tempe

^cMGH Institute of Health Professions, Boston, MA

^dUniversity of Missouri–Columbia

Correspondence to Mary Alt: malt@email.arizona.edu

[†]In memory of our colleague and collaborator, Samuel (Sam) Green, who passed away during the preparation of this article. We gratefully acknowledge his contributions to this research.

Editor-in-Chief: Sean Redmond

Editor: Jan de Jong

Received January 10, 2018

Revision received May 11, 2018

Accepted January 23, 2019

https://doi.org/10.1044/2019_JSLHR-L-18-0009

Disclosure: The authors have declared that no competing interests existed at the time of publication.

review, we specifically do not highlight the literature on non-school-age children because of the significant differences in cognitive development and environmental language demands for infants, toddlers, and adults. These inherent differences would make it difficult to use literature from other age groups and study designs to inform our predictions. In addition, there is a difference between highlighting differences in vocabulary—an end product—and understanding the mechanisms involved in word learning. Below, we review studies on school-age children that are comparable to the work and population that we use in the current study. We currently know very little about the mechanisms of word learning for school-age bilingual children, who have experience with schooling and are now learning new words, to succeed academically in more advanced subject learning (e.g., math and science concepts).

As children progress through school, differences in word learning abilities may become more pronounced, with academic repercussions inasmuch as the ability to acquire new vocabulary is an essential academic skill related to word reading and reading comprehension (e.g., Ouellette, 2006), science (e.g., Carlisle, Fleming, & Gudbrandsen, 2000), and mathematics (e.g., Riccomini, Smith, Hughes, & Fries, 2015). The kinds of words that preschool children versus school-age children are required to learn (primarily for functional communication in preschool vs. academic pursuit later) also make it difficult to generalize from preschool studies of word learning to older children. Many bilingual children in the United States are more likely to be at risk academically (Hoff, 2013) or lag behind their monolingual peers (Han, 2012), especially if they are English language learners (The Nation's Report Card, 2015). Although this finding is confounded by socioeconomic status (SES; e.g., Alt, Arizmendi, & DiLallo, 2016; Hoff, 2013), there is the potential that differences in word learning might contribute to academic challenges. It is imperative that we understand how both monolingual and bilingual school-age children learn novel vocabulary in order to (a) understand how to interpret academic findings of differences in performance between monolingual and bilingual children, (b) inform our instruction for bilingual children, and (c) understand the trajectory of bilingual language development.

In the following sections of the Introduction, we examine the logic behind possible differences between monolingual and bilingual word learning in children, the evidence for word learning differences, and the rationale behind the current study.

Why Might There Be a Difference Between Word Learning in Monolingual and Bilingual Children?

Monolingual and bilingual infants (e.g., Houston-Price et al., 2010) and toddlers (Brojde et al., 2012) approach word learning differently. Bilingual word learners are different from monolingual word learners in terms of task, environment, and potentially cognitive resources available for word learning. The fundamental task of acquiring two separate lexicons is different from the task of acquiring one. Learning

two lexicons adds cognitive load—people who are bilingual need to learn two sets of words and make constant decisions about which word to use in which situation, but as evidenced by the many successful bilingual word learners in the world, this increased load is manageable for most people. The environment of a bilingual learner is different from that of a monolingual learner. The bilingual learner routinely hears different phonology, morphology, and syntax from each language and must decide which semantic and pragmatic cues to attend to in different linguistic situations. These two factors—different task needs and environmental inputs—have led some people to posit that bilingual learners develop additional cognitive resources (e.g., superior working memory skills; Morales, Calvo, & Bialystok, 2013) to bring to the word learning process.

Word learning is a complex, dynamic process that evolves over repeated exposures (e.g., McGregor, Friedman, Reilly, & Newman, 2002). If we focus on auditory word learning or learning from listening, the task is to bind a label to a referent (e.g., “cap” = ; e.g., Alt, Plante, & Creusere, 2004). Learning a word means that one has a well-formed representation of the phonology of the label and that one could accurately produce the word or not confuse it with a similar-sounding word (e.g., “cap” vs. “gap”). It also means that one understands enough about the semantics of the referent to be able to appropriately understand the word in different contexts (e.g., realize that a cap can be something you wear or that you put on top of your tube of toothpaste, and that a cap to wear is not the same as a visor even though they both have brims).

In considering this process, there are some situations in which the differences between monolingual and bilingual word learning could emerge. When learning a word, although the referent will be the same, no matter which language a child is using for a label, a bilingual child could have two labels from which to choose for a given referent (e.g., “cap” and “gorra”). The cognitive load is higher given that the child would have to create, sort, refine, and choose the correct representation for a given situation. Although this situation is less likely to occur when learning a truly novel word, bilingual children will still have two phonological systems to choose from when trying to encode a novel word. Thus, the phonological component of word learning may be the most vulnerable to error for bilingual children. The existence of two phonological systems could potentially complicate how a child learns a novel label due to (a) potential phonological cross-linguistic influence from one language to the other, (b) lack of practice or exposure in one language compared to monolingual peers, or (c) an increased cognitive load from the need to select a given system. Below, we further explore some of the differences related to navigating two phonological systems and then discuss ramifications of differences in cognitive load that may emerge.

Cross-linguistic influences were defined by Paradis and Genesee in 1996. This term describes a phenomenon in bilingual language development in which one can discern an influence of one language on another that results in bilingual children demonstrating different patterns of language

acquisition than monolingual children. These types of influences have been widely documented for phonology and syntax. In terms of word learning, the literature is sparse.

Lack of exposure to a language could influence bilingual word learning both for receptive and expressive tasks. The simple math of bilingualism is that bilingual children have less practice in either of their languages than monolingual children, who have 100% of their language experience in a single language. Thus, when learning a novel word with English-like phonology, it may be the case that bilingual children are at a disadvantage because their relative lack of practice with English may lead to more effortful learning. Alternatively, school-age bilingual children may have sufficient practice with English to perform equivalently to peers on word learning tasks (Alt, Meyers, & Figueroa, 2013). There may be a ceiling as to how much practice is needed to be a proficient word learner in terms of phonology, and despite having measurably less exposure than monolingual peers, school-age bilingual children may have sufficient exposure to English phonology to learn as well as their monolingual peers. Currently, the sparse literature on this topic makes it difficult to predict what word learning differences in school-age children may look like.

Nearly every model of bilingual language development includes a component to account for the increased cognitive load of processing two languages compared to just one. Blumenfeld and Marian (2009) describe three separate models: (a) Li and Farkas's (2002) Self-Organizing Model of Bilingual Processing, which posits that inhibitory mechanisms are built based on language input; (b) the Bilingual Interactive Activation Model (Dijkstra & Van Heuven, 1998), which discusses inhibition between language nodes; and (c) the Inhibitory Control Model (Green, 1998), which includes a supervisory attentional system. All of these models have the same thing in common: the fact that a bilingual child would need to inhibit use and processing of one language to use the other. Thus, bilingual children may have an increased cognitive load due to the work of developing and switching between two language systems, which is work that monolingual children need not engage in. This cognitive load could also lead to vulnerabilities not only in phonology but also in other aspects of word learning, such as learning about referents and the details associated with those referents. Alternatively, the need to process additional information may lead to more efficient processing that could lead to a bilingual processing advantage. For example, this might manifest in the putative bilingual advantage for a wide range of abilities, such as cognitive control, working memory, and symbolic representation (e.g., Adesope, Lavin, Thompson, & Ungerleider, 2010). In this view, as a result of being exposed to two languages with corresponding processing demands, bilingual children may have developed linguistic processing systems that are simply more efficient than their monolingual peers. Advantages such as the ones listed above might result in a bilingual advantage in word learning. As noted earlier, there is limited information in the literature on novel word learning in school-age children who are bilingual, so this is an empirical question.

Word Learning in School-Age Bilingual Children

Only a handful of studies have evaluated novel word learning in bilingual children, and the differences in methodology make them difficult to compare. However, four outcomes could occur: (a) no difference in performance between bilingual and monolingual word learners; (b) bilingual children showing a word learning advantage over monolingual children; (c) bilingual children showing a word learning disadvantage when compared to monolingual children; or (d) bilingual children performing in a manner that is different from, but not decidedly better or worse than, monolingual children. Thus, the following studies will be described in how well they fit each of the remaining possibilities and in what areas of word learning (i.e., semantic, phonological, receptive, expressive) differences might be found.

The idea that there should be no difference between monolingual and bilingual learners finds some support in some of the results from Alt et al. (2013). These researchers evaluated cross-linguistic influences in the fast-mapping phase of word learning, when learners are first exposed to a novel word. In their study, SES-matched preschool-age and school-age Spanish–English bilingual children were asked to learn novel names for novel dinosaurs whose names had either high or low English phonotactic probability. Each age group was analyzed separately. Children were asked to decide whether a production of the newly learned name was correct or incorrect. They were also asked to produce the names of the dinosaurs. The bilingual and monolingual English children were equally sensitive to the phonotactic patterns of the English-like words. That is, all children were more accurate on high phonotactic probability nonwords than on low-probability nonwords on both naming and name identification tasks. These results showed that, for this task, bilingual children did not demonstrate evidence of cross-linguistic influences of phonotactic frequency from one language to the other. This could be taken as evidence for the no-difference hypothesis, at least as it applies to a receptive phonological task.

If bilingual children did perform poorly on a mispronunciation detection task (like the one described above), this could be for two reasons. First, the children who are bilingual could have poorer auditory perception than their monolingual peers, particularly for English phonology. Indeed, there is evidence that bilingual individuals are less accurate than monolinguals at speech perception for their second language, particularly in challenging listening situations, such as in a noisy environment (Tabri, Abou Chacra, & Pring, 2011). This would not be an inherent processing problem, but rather a relative lack of familiarity with the phonology of one language that could potentially lead to this outcome. The second possibility could be that children who are bilingual may be more open to alternate pronunciations of words (Sebastian-Galles, 2010). Although all listeners hear varied pronunciation of the same word, this is especially true for children in bilingual environments. Not only might they hear allophonic variations, but they might also

hear phonemic variations. For example, their monolingual English-speaking teacher might say “very good,” and their native Spanish-speaking aunt might produce the same phrase as “/bEri gud/,” due to the influence of Spanish on English because /v/ is not a phoneme in Spanish (e.g., Goldstein, 2004). Having regular experience with different pronunciations of the same word when the meaning is clear may make bilingual children more accepting of alternate pronunciations and less likely to label a difference in pronunciation as “wrong.” In such a situation, we would expect bilingual children to have equivalent performance to monolingual speakers on hits (when a pronunciation is correct), but lower accuracy on misses (when a pronunciation is inaccurate), because they are incorrectly accepting alternate pronunciations.

Several researchers have examined word learning in bilingual school-age children. In Alt et al. (2013), there was no evidence of a disadvantage for correctly identifying names or linking names to semantic referents for preschool or school-age bilingual children, indicating equivalent performance on receptive phonological tasks. However, despite being as sensitive to phonological patterns as their monolingual peers, bilingual preschoolers were less accurate than monolingual peers when asked to produce the learned words. This may be because the task required more cognitive resources as it required children to generate, rather than simply recognize, the phonology of the label. Although the task would require more resources for both monolingual and bilingual children, the increased load of the task on top of the increased cognitive load bilingual children already have may have led to the lower accuracy for the bilingual group. Alternatively, the deficit may have emerged because the younger children had not yet had sufficient practice with English phonology. Thus, this same study also provides evidence for the bilingual deficit hypothesis, at least for preschoolers in expressive phonological tasks.

Although the school-age children in the study of Alt et al. (2013) did not show a naming deficit or advantage, it is not clear what their performance would look like beyond the fast-mapping stage or for semantic learning because the study was not set up to measure semantic learning. In terms of the limitations of fast mapping, the average percent consonants correct for all children was 69% on the low phonotactic probability words; there was more learning to do, and differences might emerge in later stages of learning such as configuration or engagement (see Leach & Samuel, 2007).

Buac et al.’s (2016) work also provides mixed evidence, with support for both the “no-difference” and “different—not better or worse” possibilities. They studied novel word learning in school-age children (5- to 7-year-olds) to examine predictors of performance on processing tasks. The children in this study were monolingual English and bilingual English–Spanish children of the same age recruited from the same neighborhoods, although the monolingual group had a higher reported level of years of primary caregiver’s education (18.89) than the bilingual group (15.80).

Children completed a novel word learning task in which they were exposed to eight nonwords paired with either familiar (i.e., animals) or unfamiliar referents. The nonwords had English phonotactics and a CVCVC syllable shape (C = consonant, V = vowel). Children heard each word four times. In support of the no-difference possibility, the bilingual children were as accurate as their monolingual peers on a forced-choice recognition task that assessed their learning. This is another example of phonological–semantic linking being equivalent in school-age bilingual children in the early stages of word learning. The groups also performed equivalently to each other regardless of whether the referent with the label was familiar or unfamiliar, showing no difference in how they approached the semantics of the learning task.

The evidence for differences between the groups came in terms of what factors were predictive of performance. They were different for each group. Buac et al. (2016) found that the best predictor of performance for the bilingual group was English vocabulary, albeit only for word–label pairs that included familiar referents. In contrast, non-verbal intelligence was a significant predictor of word learning, but only for the monolingual group. SES did not significantly affect word learning for either group. The fact that different factors predict word learning for the groups suggests that what is contributing to their learning is different.

More support for the “different—not better or worse” outcome is found in Merriman and Kutlesic (1993). They highlighted potential differences in semantic learning processes in monolingual (English) and bilingual (Serbian–English) 5- to 8-year-olds. The children in this study were the same age and were recruited from the same neighborhood. Children were exposed to novel objects that differed in their shape, material, color, and size. When they were taught the name of the object (the number of presentations was not specified), a specific feature of the object was highlighted. Children were then asked to decide which of a set of other objects could also be referred to by the novel name. The objects were specially designed so that they varied orthogonally in terms of whether they matched the object that was directly taught in terms of having the highlighted feature (+, –) and perceptual similarity (+, –). The researchers were interested in determining if children were giving more weight to the highlighted features or the perceptual features when generalizing the newly learned term to new objects. Both groups of children were able to learn the names of the words, thus demonstrating similar phonological–semantic linking in bilingual and monolingual children. However, the older monolingual children (7- and 8-year-olds) were more likely than their bilingual age-matched peers to extend the novel name based on the highlighted feature. The point of this finding is that monolingual and bilingual children, though equally accurate, might be attending to different cues while learning words.

There is more evidence for the “different—not better or worse” position in a study by Jubenville, Sénéchal, and Malette (2014), who found that monolingual and bilingual

children were attending to different cues in a learning task. The authors examined incidental print's effect on word learning for monolingual (French) and bilingual or multilingual (French and English, and other unspecified languages) 9-year-olds. Children in this study were in the same grade, and both groups of participants came primarily from high SES homes with highly educated parents.

The authors were examining the orthographic facilitation effect, which refers to the boost people receive during oral word learning when the written (orthographic) form of the word is present during learning (e.g., Ricketts, Bishop, & Nation, 2009). Children were assigned to one of three groups: consistent print, inconsistent print, or no print. In the print conditions, the words to be learned had consistent phoneme–grapheme representations. For example, in English, a word like *cat* is consistent; the three letters of the word clearly correspond to three phonemes /k@t/. In the inconsistent condition, there was a mismatch between the phonemes and the graphemes. An English word like *late* is an example of this type of inconsistency, where there are four graphemes and only three phonemes /let/. The no-print condition did not include orthography. Children were asked to learn six words. They were presented with each word one time per block, with at least six and up to nine blocks per child. After being presented with the words, they were asked to say the names of the words. They were given feedback for incorrect productions. Thus, children had an unequal number of exposures to each word.

Both groups of children demonstrated an orthographic facilitation effect, but monolingual children showed a stronger orthographic facilitation effect when the print had a consistent phoneme–grapheme relationship. In contrast, the bilingual children learned more when the print had an inconsistent phoneme–grapheme relationship—in this case, usually a silent consonant at the end of the word. The authors interpret these differences as potentially due to differences in how children allocated their attention during training, either paying more attention to the inconsistent words or utilizing potentially superior metalinguistic skills. This, such as the Merriman and Kutlesic (1993) results, highlights the fact that potential differences in word learning in bilingual children may not manifest in overall accuracy, but rather in a profile of accuracy across different measures. Although Merriman and Kutlesic demonstrated this with semantic cues, Jubenville et al. (2014) demonstrate it with orthography and phonology.

What we know about novel word learning in bilingual children is still limited. Practically, we need to examine word learning beyond the fast-mapping stage, as differences between monolingual and bilingual children may emerge beyond the first few exposures. We also need to determine which types of cues bilingual children may attend to as we know that monolingual and bilingual children, though often equally accurate, may attend to different cues while learning words. No study, to date, has asked children to learn, recognize, and produce both phonological and semantic components of novel words, which limits our ability to know

which cues might take precedence for each group. Although there are theoretical reasons to suspect a bilingual word learning advantage, to date, no study has provided evidence for this possibility. The inclusion of a study that has a more demanding task (e.g., requiring children to attend to more elements of the word/reference pair) might have the potential to reveal such an advantage.

The Current Study

To answer our research question, we tested whether bilingual (Spanish–English) second graders with typical development differed from their monolingual (English) peers with typical development on tasks in which children needed to perceive and store the phonology of a label and the characteristics (semantics) of a referent and then link this phonological and semantic information. By including different aspects of word learning in different tasks, we will be able to be clear on which aspect of word learning, if any, may be different for bilingual children. Ours will be one of the first studies to examine word learning in bilingual school-age children across more than just a few exposures to a word. This will give us a better sense of whether any of the differences noted in the literature manifest beyond initial exposure. Because the literature would suggest that differences could manifest in differing profiles of performance, we included different types of manipulations in our word learning paradigms (e.g., word length, visual similarity). Based on the literature, many options are possible, and we are open to all of them. However, the existing evidence points most strongly to the following hypotheses:

1. We expect to find no differences between groups on phonological–visual linking tasks.
2. We may find a bilingual disadvantage for tasks that require the production of phonological labels, but not those that require the recognition of phonological labels.
 - a. However, if there is a disadvantage for recognition of phonological labels, it might be due to openness to alternate interpretations.
3. We may find differences in which cues bilingual and monolingual children attend to during the word learning tasks.

The nature of which cues different children may attend to is the most exploratory of the analyses, given the difference in design between the few studies on novel word learning in school-age bilingual children. There is the least information on semantic cues, and as described above, the outcomes could vary.

Method

Participants

Participants were monolingual (English; $n = 167$) and bilingual (Spanish–English; $n = 76$) second graders with

typical development.¹ All children were between 7 and 9 years of age and were recruited through schools and community centers in Arizona, Massachusetts, and Nebraska after we received institutional review board approval for the study. Table 1 includes the race and ethnicities of both groups, as reported by their parents. To be considered typically developing, the following criteria had to be met (see Table 2).

To be included in the monolingual English group, parents had to report their child's primary language as English, his or her primary caregivers had to speak only English, and his or her schooling (prekindergarten to second grade) must have been only in English. Bilingual children could have English, Spanish, or both reported as their primary language(s). They needed to have at least one primary caregiver who spoke Spanish to the child. Their preschool education (if they attended) needed to complement their home environment. That is, if there was no English in their home environment, their preschool experience needed to contain English or both English and Spanish. If there was some English in the home, their preschool experience could be in English, Spanish, or both. Kindergarten through second grade instruction needed to be in English or both Spanish and English. Finally, parents needed to report that children could carry on a conversation in both English and Spanish. Thus, all participants had functional skills in both languages.

Figure 1 illustrates our process for determining whether a bilingual child had typical language skills. It includes the use of both the English (Semel, Wiig, & Secord, 2003) and Spanish (Semel, Wiig, & Secord, 2006) versions of the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4). This allowed us to assess children's language skills in each of their languages, but neither test was normed on bilingual children. Therefore, to determine whether a child was language impaired, we used data-based cutoffs for the CELF-4 Spanish from Barragan, Castilla-Earls, Martinez-Nieto, Restrepo, and Gray (2018). They analyzed data from the CELF-4 Spanish from over 650 bilingual children in the Phoenix metropolitan area and found that a standard score cutoff of 78 had good sensitivity (86%) and specificity (80%) for this population.

¹Participants in this article represent a portion of the participants in a larger sample from the POWWER study, funded by National Institute of Health NIDCD Grant R01 DC010784. The POWWER study includes the groups reported, as well as children with language impairment, children with dyslexia, and children with comorbid dyslexia and language impairment. Participants in the POWWER study completed a total of six word learning games and a comprehensive battery of working memory tasks (see Cabbage et al., 2017), completed over the course of at least 6 days. A portion of the data for the monolingual children in this study was reported in Alt et al. (2017), and a portion from both the monolingual and bilingual groups was reported in Erikson et al. (2018). Data from the POWWER data set have also been published in Arizmendi et al. (2018), Baron et al. (2018), Cowan et al. (2017), Gray et al. (2017), and Green et al. (2016).

Individuals who are bilingual can be characterized in many ways. Given our potential concerns about adequate exposure to English, we were aiming for a more balanced bilingual sample, meaning that children demonstrated relatively strong skills in both English and Spanish. The majority of our bilingual participants could be defined as simultaneous bilinguals, that is, they began learning both Spanish and English since birth or very early in life, with 93% of families reporting both English and Spanish being spoken in the home either by the parents (63%) or other family members (e.g., siblings, grandparents; 30%). Three families (3.9%) reported only speaking Spanish in the home. One of these children received all schooling in English, whereas the other two reportedly had a mixture of Spanish and English education. Two families (2.56%) in the sample declined to provide information about the languages spoken in their home.

In addition to our inclusionary measures, we also report descriptive measures of English vocabulary, paragraph comprehension, parental ratings of attention, and, for the bilingual group, measures of Spanish vocabulary. Table 3 includes the demographics of both groups. Table 4 includes the data on the Spanish-only measures.

Stimuli

Our experiment consisted of six different word learning manipulations (games) each of which assessed learning using the same five tasks. Those tasks measured phonological–visual linking (phonology, semantics), naming (phonology), mispronunciation detection (phonology), visual feature recall (semantics), and visual difference decision (semantics). The tasks will be described in more detail below, but see Figure 2 for an overview. Each game had a different focus, including word length comparisons (two- vs. four-syllable labels), phonological similarity (similar or dissimilar labels), visual similarity (similar or dissimilar referents), location (referent appears in stable or variable location), orthography (orthography present or absent during learning), and verbs, in which children were to learn the name of the action rather than the name of the referent.

Each game took roughly 30 min to play, and children played only one game each day to decrease the possibility of interference effects. Games were administered on a touch-screen computer in the context of a pirate adventure in which children earned virtual coins for correct answers. These could then be redeemed at the virtual pirate store.

Each game had four blocks. After the first block ended, the phonological–visual linking task resumed. In Blocks 2 through 4, children had 15 attempts to match each monster with its name. Thus, by the end of each game, there were 47 attempts to match each monster with its name ($2 + 15 + 15 + 15$). After each block, the other tasks were introduced, always in random order. Thus, at the end of the game, each child was prompted to name each monster or action four times, made decisions about the name and appearance of each monster or its actions four

Table 1. Distribution of parentally reported ethnicity and race by group.

Group	Total N	Ethnicity			Race					Not reported
		Hispanic	Not Hispanic	Not reported	White	More than one race	Black/African American	American Indian/ Alaska Native	Asian	
Monolingual	167	20	145	2	135	20	4	3	3	2
Bilingual	76	76	1	1	41	4	1	2	0	28

Note. There are a high number of “not reported” not because data were not collected but because many Hispanic participants choose not to select a race.

times, and recreated the appearance of each monster or its action four times.

Games

The main point of the games was to learn the names and characteristics of novel sea monsters who inhabited a virtual pirate world. Each game followed the same structure. They began with the phonological–visual linking task. In this task, children saw four different monsters on a touch screen and heard a name (or action for the verb game). For the verb game, children saw the same monster, but performing four different novel actions. For the verb game, they heard the action named as an uninflected imperative, akin to “dance!” They had to touch the monster that they thought matched the name or the action, which required them to link phonological and semantic representations. They were given immediate feedback in the form of a coin for a correct answer or a rock for an incorrect answer. These virtual tokens appeared on the bottom of the screen with their accompanying sounds (clinking coin, thudding rock) immediately following the child’s choice (see Figure 3). In the first block of the game, children had two attempts to correctly link the phonological (label) and visual (referent) information for each of the four monsters. Then, children were presented with the remaining four tasks, in random order. The second, third,

and fourth blocks followed, with 15 exposures to the label of each monster per block.

In the naming task, children saw a monster and were asked to state its name or the name of the action it was performing in the verb game. This required that they create and store a detailed phonological representation of the name. They were given positive feedback for attempting to name the monster, but no specific feedback about the accuracy of their response. They named each monster, in random order, one time per block.

In the mispronunciation detection task, children saw a monster and heard either its correct name (or the name of its action) or a phonologically related foil. The foils differed from the correct name by a manipulation of the final consonant (e.g., nudwef/nudweg/nudwev). Children had to press a button on the screen, indicating whether the name they heard was correct or incorrect. Thus, they needed to access the detailed phonological representation they had created to make this decision. They made one decision for each monster per block and were provided with immediate feedback using the coin/rock paradigm.

In the visual difference decision task, children saw a monster that was either an accurate representation of that monster (or its action) or a semantically related foil. The foils varied from the accurate referent by one to three semantic features (e.g., color, head covering, eyes). In the

Table 2. Inclusionary criteria for monolingual and bilingual groups.

Measure	Monolingual English	Bilingual Spanish–English
Family questionnaire	7–9 years old in 2nd grade No history of neuropsychiatric disorders (e.g., ADHD, autism) No history of special education services	
Child has repeated a grade	Excluded	
Vision: near acuity	Binocular only at 20/32 (or 20/40 in bad lighting conditions, must be documented)	
Vision: color	Correct response on 8 or 9 color plates	
Hearing screening	20 dB at 1, 2, and 4 Hz (or 30 dB in noisy conditions, must be documented)	
GFTA-2	≥ 31st percentile (if < 31st percentile, study phonemes must be produced correctly in 2/3 production attempts)	
K-ABC2 Nonverbal standard score	≥ 75	
TOWRE-2 Composite standard score for grade	≥ 96	
CELF-4 Core Language standard score	≥ 88	(see Figure 2)

Note. ADHD = attention-deficit/hyperactivity disorder; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); K-ABC2 = Kaufman Assessment Battery for Children, Second Edition (Kaufman & Kaufman, 2004); TOWRE-2 = Test of Word Reading Efficiency–Second Edition (Torgesen, Wagner, & Rashotte, 2012); CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel, Wiig, & Secord, 2003).

Figure 1. Process for determining whether a bilingual child had typical language skills. CELF = Clinical Evaluation of Language Fundamentals; DQ = disqualify; FS = formulated sentences.

verb game, the foils related to the speed of the action (e.g., fast or slow), the direction of the action (e.g., horizontal, circular, random), the type of action (e.g., spinning, stretching), and any special effects (e.g., color changes). Children had to press a button on the screen, indicating whether the monster or the action they saw was accurate or not. This required that they access a detailed semantic representation of the monster or action that they had created during the phonological–visual linking task. They made one decision for each monster per block and were provided with immediate feedback using the coin/rock paradigm for each monster.

In the visual feature recall task, children were presented with the outline of a monster and four choices for each of the following semantic features: color, eyes, arms, and head covering (see Figure 4). For the verb game, the four semantic features were speed, direction, type of action, and special effects (see Figure 5). For the verb task, the choices were animated in order to demonstrate these different characteristics. The participants’

job was to accurately recreate the semantic features of the monster or action, which required them to access the detailed semantic representation of the monster or action they had created. Once they selected a feature, it would fill in on the shape so they could “try out” the look. Once they were confident in their selection, they pressed an “I’m done” button and were immediately given feedback for each feature.

Referent Images

The referents were sea monsters created especially for this task. The referents were made to be the same size. Each game used a unique set of referents.

Labels

Each game presented unique labels for unique sea monsters. All of the labels were nonwords with a CVC–CVC structure composed of English phonemes (e.g., /gompæv/). All of the labels had low phonotactic probability in English and no neighbors. The only exception to this

Table 3. Means and standard deviations for standard scores on inclusionary and descriptive assessments.

Variable	Monolingual	Bilingual	BF ₁₀ ^a	Hypothesis supported
<i>n</i>	167	76	<i>n/a</i>	
Age	7;6 (0;4)	7;9 (0;5)	27.74	Alternative
MLE	15.38 (1.65)	12.59 (2.57)	3.717e + 16	Alternative
TOWRE-2	109.44 (8.40)	108.05 (7.90)	0.30	Null
K-ABC2	117.60 (15.52)	106.39 (11.75)	196585.63	Alternative
CELF-4	108.75 (9.58)	93.76 (9.23)	1.548e + 21	Alternative
GFTA-2 ^b	50.89 (8.53)	44.50 (10.83)	10446.09	Alternative
EVT-2	112.38 (10.95)	93.81 (9.09)	4.580e + 25	Alternative
WRMT	108.22 (9.85)	102.34 (9.20)	1067.13	Alternative
ADHD	10.19 (8.76)	7.83 (8.02)	0.82	Anecdotal (Null)

Note. Lower scores on this measure reflect fewer concerns. MLE = mother's level of education; TOWRE-2 = Test of Word Reading Efficiency—Second Edition (Torgesen et al., 2012); K-ABC2 = Kaufman Assessment Battery for Children, Second Edition (Kaufman & Kaufman, 2004); CELF-4 = Clinical Evaluation of Language Fundamentals—Fourth Edition (Semel, Wiig, & Secord, 2003); GFTA-2 = Goldman-Fristoe Test of Articulation—Second Edition (Goldman & Fristoe, 2000); EVT-2 = Expressive Vocabulary Test—Second Edition (Williams, 2007); WRMT = Woodcock Reading Mastery Test, Paragraph Comprehension Subtest (Woodcock, 2011); ADHD = parental rating of attention-deficit/hyperactivity disorder (ADHD) behaviors using the ADHD Rating Scale—IV Home Version (DuPaul et al., 1998).

^aBetween-groups differences were tested using Bayesian independent-samples *t* tests. If the alternative hypothesis is supported, it means there is evidence for a group difference. If the null hypothesis is supported, it means there is evidence for no group difference. ^bPercentile, rather than standard score.

characterization is that, in the game that contrasted word length, there were two 4-syllable labels. None of these nonwords had syllables that held lexical content in Spanish. A list of the labels is in Appendix A.

Analytic Approach

We analyzed results from each task within each game separately using Bayesian repeated-measures analyses of covariance (ANCOVAs), with nonverbal intelligence as the covariate, using JASP (Version 0.8.6; JASP Team, 2018). Bayesian analysis is robust to group size differences and measures the probability of hypotheses in either direction: the null or a reasonable set of alternative hypotheses (Kruschke, 2013). In our case, we were examining potential group differences; thus, the null hypothesis would

be no group differences and the alternative hypothesis would be group differences. Bayesian analysis, instead of providing an *F* statistic for each main effect and interaction, tests the “relative predictive success of two or more models” (Etz, Gronau, Dablander, Edelsbrunner, & Baribault, 2018, p. 224) and provides a Bayes factor for each model (e.g., main effects, interactions). The interpretation of the strength of evidence for any model is drawn from guidance provided in Wagenmakers et al. (2018). In this scheme, when the Bayes factor is stated in terms of the likelihood of the alternative hypothesis divided by the likelihood of the null, a Bayes factor of 10 or above is considered strong evidence for the alternative hypothesis, a Bayes factor of 3 or above is considered as moderate evidence for it, and a Bayes factor of above 1 but less than 3 is considered as anecdotal (essentially indeterminate) evidence. Expressed this same way, moderate evidence for the null is 1/3 or less, and strong evidence for the null is 1/10 or less, that is, the reciprocal of the criteria for the alternative hypothesis. The scale includes guidelines for even stronger evidence such as very strong (30–100 or 1/100–1/30) and extreme (> 100 or < 1/100).

For all tasks, group (monolingual, bilingual) was the between-groups factor. Manipulation (e.g., long vs. short words) was the within-group factor, with the exception of the verb game, which had no manipulations. An additional within-group factor was block (1, 2, 3, 4) for the phonological–visual linking, naming, and visual feature recall tasks. After an initial analysis, we found that there were no significant Group × Block interactions; therefore, we omitted block from the analyses presented below. This allowed us to better gauge the effect of the covariate. Within JASP, one can compare matched models to find the relative strength of each factor. However, this analysis excludes higher order interactions (Wagenmakers

Table 4. Means, standard deviations, and ranges for Spanish-only assessments.

Measure	<i>n</i>	<i>M</i> (<i>SD</i>)	Range
SCELF total ^a	23	95.08 (11.17)	77–117
SCELF-FO	74	10.85 (2.32)	6–16
EOWPVT-4: SBE	74	110.20 (13.91)	79–145
EOWPVT-4: SBE ratio	74	0.52 (0.21)	0.06–0.93

Note. SCELF total = Spanish Clinical Evaluations of Language Fundamentals standard score (Semel et al., 2006); SCELF-FO = Spanish Clinical Evaluations of Language Fundamentals—Formulation Oraciones standard score; EOWPVT-4: SBE = Expressive One-Word Picture Vocabulary Test—4: Spanish-Bilingual Edition standard score (Brownell, 2001); EOWPVT-4: SBE ratio = Expressive One-Word Picture Vocabulary Test—4: Spanish-Bilingual Edition total raw Spanish words produced/total raw words produced.

^aNot all children needed to take the entire SCELF. Please refer to Figure 1 for the inclusion decision process for bilingual children.

Figure 2. Game details and task details.

et al., 2018). By excluding block, we were able to determine the strength of any group effects separate from the effect of nonverbal intelligence.

We used independent-samples Bayesian *t* tests to clarify any interactions. In traditional statistics, when one runs multiple post hoc tests, there is a risk of Type II error

Figure 3. Example of the phonological–visual linking task where the child makes a correct response.

or one of them being significant just by chance. In the Bayesian framework, each *t* test yields the best estimate of the strength of the alternative hypothesis compared to the strength of the null. Each one can go either way, so there is no special adjustment needed for multiple tests (Etz et al., 2018).

Our decision to covary nonverbal intelligence was based on our expectation of no between-groups differences on this measure. However, as Table 3 shows, there was a significant between-groups difference on nonverbal intelligence (Kaufman Assessment Battery for Children, Second Edition; Kaufman & Kaufman, 2004); therefore, we used it as a covariate. There were expected between-groups differences on other measures (e.g., scores on English-only measures, SES). There are conflicting findings regarding the relationship between SES and word learning. For example, Calvo and Bialystok (2014) found an effect of SES for 6-year-old bilingual children, whereas Buac et al. (2016) found that SES did not predict word learning for the bilingual children in their study. Thus, in our initial analyses, we covaried maternal level of education, our proxy for SES; however, it was not significantly related to any of our outcome measures. Thus, we used only nonverbal intelligence as a covariate in the reported analyses.

Figure 4. Example of the visual feature recall task before the child has made a response.

Results

Detailed results from the repeated-measures ANCOVAs for all tasks may be found in Appendix B.

Hypothesis 1 (Phonological–Visual Linking)

We expected no differences between groups on phonological–visual linking tasks. We examined performance on the phonological–visual linking task for each of our six games. These games manipulated either word length, phonological similarity, visual similarity, location of the referent, or orthography for nouns, and one examined verb learning (see Table 5). Results of the repeated-measures ANCOVA suggested no difference between the groups on

Figure 5. Example of the visual feature recall task for the verb game. Although these are still images, the actual task was animated to better illustrate these concepts.

five of the games. The evidence was moderate for the location and verb games and anecdotal for the word length, visual similarity, and orthography games. There was no more than anecdotal evidence for a between-groups difference with an interaction between group and condition for the phonological similarity game. Follow-up *t* tests for the interaction showed moderate evidence for the null hypothesis of no between-groups differences for both the similar and dissimilar conditions. Thus, for all the games, there was evidence of no difference between the groups, although that evidence was only anecdotal for three games.

Hypothesis 2 (Naming Versus Mispronunciation Detection)

We expected that we might find a bilingual disadvantage for tasks that required the production of phonological labels (i.e., the naming task), but not those that only required the recognition of phonological labels (i.e., the mispronunciation detection task).

Naming

The naming task included the same manipulations (e.g., short and long nonwords) as the phonological–visual linking task. However, in this task, children had to produce the name of each monster. Details for the results of all the naming games are in Table 6. There was no more than anecdotal evidence for between-groups differences for the phonological similarity and location games for an interaction between group and condition. Follow-up *t* tests for the phonological similarity game provided only anecdotal evidence of a group difference in the similar condition (favoring the monolingual group) and moderate evidence for the null hypothesis for the dissimilar condition. For the location game, there was moderate evidence favoring the monolingual group for the stationary condition, as opposed to moderate evidence for no group differences for the variable condition (see Figure 6). The ANCOVA showed moderate evidence for the null hypothesis of no group differences for the word length, visual similarity, orthography, and verb games.

Mispronunciation Detection

This task required participants to decide if a label that they heard was correct or not. Table 7 provides detailed results for this task. We found strong evidence for between-groups differences favoring the monolingual group on three games that had heavier visual demands: visual similarity, location, and orthography. For the remaining three games (word length, phonological similarity, verbs), there was only anecdotal evidence for the null hypothesis of no group differences.

To determine if the bilingual children were being more lenient in their judgments of correct pronunciations, we examined the results from the mispronunciation task in terms of “hits” (correct affirmative responses) and “foils” (correct negative responses). If there was a pattern of bilingual children responding with more “yes” answers, we would expect an interaction on this test, with equivalent

Table 5. Mean percentage of accuracy (standard error) across all four blocks of the phonological–visual linking task with Bayes inclusion factors and interpretations with nonverbal intelligence as a covariate.

Word length	Phonological similarity	Visual similarity	Location	Orthography	Verbs
Monolingual					
69.4 (1.2)	Similar: 60.5 (1.5) Dissimilar: 61.8 (1.5)	62.6 (1.4)	63.0 (1.4)	69.8 (1.4)	53.7 (1.5)
<i>n</i> = 159	<i>n</i> = 162	<i>n</i> = 162	<i>n</i> = 160	<i>n</i> = 130	<i>n</i> = 155
Bilingual					
68.7 (1.9)	Similar: 61.8 (2.2) Dissimilar: 60.7 (2.1)	61.6 (2.1)	60.3 (1.9)	71.1 (2.2)	50.4 (2.0)
<i>n</i> = 76	<i>n</i> = 75	<i>n</i> = 75	<i>n</i> = 76	<i>n</i> = 71	<i>n</i> = 73
Bayes inclusionary factor					
Groups tend to not differ (BF _{INC} = 0.35, anecdotal)	Groups tend to not differ (BF _{INC} = 0.51, anecdotal)	Groups tend to not differ (BF _{INC} = 0.36, anecdotal)	Groups do not differ (BF _{INC} = 0.31, moderate)	Groups tend to not differ (BF _{INC} = 0.98, anecdotal)	Groups do not differ (BF _{INC} = 0.18, moderate)
Interaction is null (BF _{INC} = 0.19, moderate)	^a Groups tend to differ in interaction (BF _{INC} = 2.14, anecdotal)	Interaction is null (BF _{INC} = 0.16, moderate)	Interaction is null (BF _{INC} = 0.16, moderate)	Interaction is null (BF _{INC} = 0.17, moderate)	No interactions possible in this game

Note. We checked for interactions but reported the findings for the strongest model that included group. The Bayes inclusion factor “...compares models that contain the effect to the equivalent models stripped of the effect.” (JASP 0.8.6), an analysis suggested by S. Mathôt (Wagenmakers et al., 2018).

^aPost hoc tests showed moderate evidence for no differences between the groups in both the similar (BF₁₀ = 0.16) and dissimilar (BF₁₀ = 0.17) conditions.

Table 6. Mean percentage accuracy (standard error) across all four blocks of the naming task with Bayes inclusion factors and interpretations with nonverbal intelligence as a covariate.

Word length	Phonological similarity	Visual similarity	Location	Orthography	Verbs
Monolingual					
28.4 (1.0)	Similar: 28.6 (1.4) Dissimilar: 27.4 (1.4)	25.6 (1.0)	Stationary: 32.0 (1.4) Variable: 28.4 (1.4)	35.1 (1.4)	19.8 (1.0)
<i>n</i> = 154	<i>n</i> = 158	<i>n</i> = 159	<i>n</i> = 157	<i>n</i> = 129	<i>n</i> = 150
Bilingual					
26.0 (1.1)	Similar: 23.3 (1.8) Dissimilar: 26.7 (1.7)	23.5 (1.2)	Stationary: 24.9 (1.6) Variable: 26.8 (1.7)	31.7 (1.9)	16.0
<i>n</i> = 70	<i>n</i> = 63	<i>n</i> = 69	<i>n</i> = 70	<i>n</i> = 65	<i>n</i> = 64
Bayes inclusionary factor					
Groups do not differ (BF _{INC} = 0.19, moderate)	Groups tend to not differ (BF _{INC} = 0.34, anecdotal)	Groups do not differ (BF _{INC} = 0.15, moderate)	Groups tend to not differ (BF _{INC} = 0.54, anecdotal)	Groups do not differ (BF _{INC} = 0.19, moderate)	Groups do not differ (BF _{INC} = 0.32, moderate)
Interaction suggests null hypothesis (BF _{INC} = 0.47, anecdotal)	^a Groups tend to differ in interaction (BF _{INC} = 1.07, anecdotal)	Interaction is null (BF _{INC} = 0.19, moderate)	^b Groups tend to differ in interaction (BF _{INC} = 1.16, anecdotal)	Interaction is null (BF _{INC} = 0.16, moderate)	No interactions possible in this game

Note. We checked for interactions but reported the findings for the strongest model that included group. The Bayes inclusion factor “...compares models that contain the effect to the equivalent models stripped of the effect.” (JASP 0.8.6), an analysis suggested by S. Mathôt (Wagenmakers et al., 2018).

^aPost hoc tests showed anecdotal evidence for group differences for the similar condition (BF₁₀ = 1.32) and moderate evidence for the null hypothesis for the dissimilar (BF₁₀ = 0.16) condition. ^bPost hoc tests showed moderate evidence for group differences for the stationary condition (BF₁₀ = 9.64) and moderate evidence for the null hypothesis for the variable (BF₁₀ = 0.19) condition.

Figure 6. Means and standard errors for both groups by condition averaged across all four learning blocks for the location naming game. The asterisk indicates a condition that had moderate evidence for the alternative hypothesis.

performance on hits, but a monolingual advantage for foils. Using a Bayesian repeated-measures ANCOVA with non-verbal intelligence as the covariate, we indeed found evidence for a group difference for the predicted interaction. The evidence for the interaction was only anecdotal for two games (word length $BF_{Inc} = 1.75$, phonological similarity $BF_{Inc} = 2.09$), but it was moderate for one game (visual similarity $BF_{Inc} = 4.21$) and very strong for two games

(location $BF_{Inc} = 33.35$, orthography $BF_{Inc} = 38.64$). In contrast, there was moderate evidence for the null hypothesis regarding the interaction for one game (verbs $BF_{Inc} = 0.21$). In the games that had evidence for the interaction, we used Bayesian independent t tests to clarify the interactions. As predicted, all five of these games had moderate evidence (word length $BF_{10} = 7.43$), strong evidence (phonological similarity $BF_{10} = 13.61$), or extreme evidence

Table 7. Mean adjusted score (standard error) on the mispronunciation detection task with Bayes inclusion factors and interpretations with nonverbal intelligence as a covariate.

Word length	Phonological similarity	Visual similarity	Location	Orthography	Verbs
Monolingual					
49.1 (2.3) <i>n</i> = 159	32.3 (2.1) <i>n</i> = 162	45.1 (2.1) <i>n</i> = 162	44.4 (2.5) <i>n</i> = 160	49.6 (2.7) <i>n</i> = 132	38.7 (2.5) <i>n</i> = 155
Bilingual					
39.0 (3.6) <i>n</i> = 76	21.8 (2.9) <i>n</i> = 75	28.1 (3.5) <i>n</i> = 76	28.8 (3.3) <i>n</i> = 76	34.5 (3.4) <i>n</i> = 71	30.0 (3.4) <i>n</i> = 73
Bayes inclusionary factor					
Groups tend to not differ ($BF_{Inc} = 0.34$, anecdotal)	Groups tend to not differ ($BF_{Inc} = 0.65$, anecdotal)	Groups differ ($BF_{Inc} = 18.53$, strong)	Groups differ ($BF_{Inc} = 10.36$, strong)	Groups differ ($BF_{Inc} = 12.55$, strong)	Groups tend to not differ ($BF_{Inc} = 0.41$, anecdotal)
Interaction is null ($BF_{Inc} = 0.20$, moderate)	Interaction is null ($BF_{Inc} = 0.17$, moderate)	Interaction suggests null hypothesis ($BF_{Inc} = 0.45$, anecdotal)	Interaction is null ($BF_{Inc} = 0.23$, moderate)	Interaction is null ($BF_{Inc} = 0.19$, moderate)	No interactions possible in this game

Note. We checked for interactions but reported the findings for the strongest model that included group. The Bayes inclusion factor "...compares models that contain the effect to the equivalent models stripped of the effect." (JASP 0.8.6), an analysis suggested by S. Mathôt (Wagenmakers et al., 2018).

(orthography $BF_{10} = 163.63$, location $BF_{10} = 360.89$, visual similarity $BF_{10} = 876.60$) in support of a between-groups difference favoring the monolingual group on foils. This contrasted with the results for hits, which ranged from anecdotal (visual similarity $BF_{10} = 0.45$) to moderate (location $BF_{10} = 0.17$, word length $BF_{10} = 0.15$, phonological similarity $BF_{10} = 0.15$, orthography $BF_{10} = 0.15$) evidence for the null hypothesis. Please see Figure 7.

Hypothesis 3 (Cues in Visual Difference Detection and Visual Feature Recall)

We expected that we might find differences in which cues bilingual and monolingual children attended to during the word learning tasks.

Visual Difference Decision Task

In this task, participants needed to make a decision about the accuracy of a visual representation of a monster. Table 8 has details for these analyses. There was only anecdotal evidence in favor of the bilingual group for the location game. There was also only anecdotal evidence for an interaction between group and condition for the orthography game. However, Bayesian independent-group *t* tests to explore the interaction showed very strong evidence that the monolingual group was more accurate than the bilingual group for the orthography present condition versus moderate evidence that there was no difference between groups in the no orthography condition (see Figure 8).

The word length, phonological similarity, visual similarity, and verb games all had moderate evidence for the null hypothesis of no difference between the groups.

Visual Feature Recall Task

In this task, participants attempted to correctly choose four semantic features of a monster. The details for the analyses of this task can be found in Table 9. There was anecdotal evidence for a group difference favoring the bilingual group for the location and verb games. The phonological similarity, visual similarity, and orthography games had moderate evidence for no group differences, whereas the word length game had only anecdotal evidence for the null hypothesis of no difference between groups. Table 10 summarizes the results for each task, game, and condition.

Discussion

Our research question was whether we would find word learning differences in typically developing monolingual and bilingual children on novel word learning tasks that used English phonology. The literature allowed for contrasting outcomes, including a bilingual word learning advantage due to more efficient processing (e.g., Mattock, Polka, Rvachew, & Krehm, 2010), a monolingual advantage due to either less practice in English or the higher cognitive demands of bilingual word learning (e.g., Alt et al., 2013; production task), or similar accuracy with evidence

Figure 7. Percent correct and standard error for hits and foils on the mispronunciation detection task by group for all games. Asterisks indicate conditions that had at least moderate evidence for the alternative hypothesis.

Table 8. Average mean percentage accuracy (standard error) on the visual difference decision task with Bayes inclusion factors and interpretations with nonverbal intelligence as a covariate.

Word length	Phonological similarity	Visual similarity	Location	Orthography ^a	Verbs
Monolingual					
68.2 (2.0)	67.0 (1.3)	77.4 (1.5)	61.8 (1.8)	– Ortho: 79.0 (2.1) +Ortho: 81.4 (2.2)	52.0 (2.1)
<i>n</i> = 159	<i>n</i> = 162	<i>n</i> = 162	<i>n</i> = 160	<i>n</i> = 132	<i>n</i> = 155
Bilingual					
67.4 (3.1)	65.5 (1.8)	76.5 (2.4)	65.1 (2.1)	– Ortho: 74.6 (4.2) +Ortho: 66.5 (4.6)	44.3 (3.9)
<i>n</i> = 76	<i>n</i> = 75	<i>n</i> = 76	<i>n</i> = 76	<i>n</i> = 71	<i>n</i> = 73
Bayes inclusionary factor					
Groups do not differ (BF _{INC} = 0.25, moderate)	Groups do not differ (BF _{INC} = 0.27, moderate)	Groups do not differ (BF _{INC} = 0.16, moderate)	Groups tend to differ (BF _{INC} = 1.27, anecdotal)	Groups tend to not differ (BF _{INC} = 0.41, anecdotal)	Groups do not differ (BF _{INC} = 0.18, moderate)
Interaction suggests null hypothesis (BF _{INC} = 0.35, anecdotal)	Interaction suggests null hypothesis (BF _{INC} = 0.54, anecdotal)	Interaction is null (BF _{INC} = 0.16, moderate)	Interaction is null (BF _{INC} = 0.17, moderate)	^b Groups tend to differ in interaction (BF _{INC} = 2.29, anecdotal)	No interactions possible in this game

Note. We checked for interactions but reported the findings for the strongest model that included group. The Bayes inclusion factor “...compares models that contain the effect to the equivalent models stripped of the effect.” (JASP 0.8.6), an analysis suggested by S. Mathôt (Wagenmakers et al., 2018).

^a– Ortho is the condition where orthography is not present, and + Ortho is the condition where orthography was present. ^bPost hoc tests showed very strong evidence for a group difference in the orthography present condition (BF₁₀ = 46.09) and moderate evidence for no group difference in the orthography absent condition (BF₁₀ = 0.26) conditions.

Figure 8. Means and standard errors for both groups averaged across all four blocks for the visual difference task for the orthography game. The asterisk indicates a condition that had very strong evidence for the alternative hypothesis.

Table 9. Mean percentage accuracy (standard error) across all four blocks of the visual feature recall task with Bayes inclusion factors and interpretations with nonverbal intelligence as a covariate.

Word length	Phonological similarity	Visual similarity	Location	Orthography	Verbs
Monolingual					
64.0 (1.2) <i>n</i> = 159	67.0 (1.3) <i>n</i> = 162	68.4 (1.1) <i>n</i> = 162	68.6 (1.2) <i>n</i> = 160	63.6 (1.2) <i>n</i> = 130	44.0 (1.3) <i>n</i> = 155
Bilingual					
63.1 (1.7) <i>n</i> = 76	65.9 (1.8) <i>n</i> = 75	65.4 (1.6) <i>n</i> = 76	69.6 (1.5) <i>n</i> = 76	62.7 (1.6) <i>n</i> = 71	45.8 (1.4) <i>n</i> = 73
Bayes inclusionary factor					
Groups tend to not differ (BF _{INC} = 0.34, anecdotal) Interaction is null (BF _{INC} = 0.19, moderate)	Groups do not differ (BF _{INC} = 0.33, moderate) Interaction is null (BF _{INC} = 0.24, moderate)	Groups do not differ (BF _{INC} = 0.18, moderate) Interaction is null (BF _{INC} = 0.29, moderate)	Groups tend to differ (BF _{INC} = 1.06, anecdotal) Interaction is null (BF _{INC} = 0.15, moderate)	Groups do not differ (BF _{INC} = 0.20, moderate) Interaction is null (BF _{INC} = 0.17, moderate)	Groups tend to differ (BF _{INC} = 1.31, anecdotal) No interactions possible in this game

Note. We checked for interactions but reported the findings for the strongest model that included group. The Bayes inclusion factor "...compares models that contain the effect to the equivalent models stripped of the effect." (JASP 0.8.6), an analysis suggested by S. Mathôt (Wagenmakers et al., 2018).

of each group focusing on different cues. Our findings are one of the few to examine novel word learning so comprehensively, beyond the fast-mapping stage, and in multiple contexts using multiple assessments of both phonology and semantics.

Our findings (summarized in Table 10) more strongly supported the hypothesis of similar performance on at least half of our tasks ($N = 15$), with clear monolingual advantages for five tasks and anecdotal or inconclusive evidence for 10 tasks. When interpreted in context, these results may best be described as supporting the literature that finds equivalent accuracy in word learning, but with differences in focus. To begin, there were no between-groups differences in accuracy on the task that asked children to link phonological (labels) and visual (referents) information. Thus, when it comes to linking novel labels and referents, there seem to be no obvious differences between typically developing monolingual and bilingual school-age learners in our study. This finding is consistent with the limited literature on novel bilingual word learning and supports Hypothesis 1.

When we turn to tasks that tax phonology, we did see some differences between the groups in terms of accuracy, in favor of the monolingual group. However, it is important to note that these differences were only apparent in particular situations. For the naming task, there was one clear instance of between-groups difference favoring the monolingual group. This contrasts with the findings from Alt et al. (2013), who found that bilingual 7-year-olds were as accurate as their monolingual peers during a naming fast-mapping task. However, this study extends the naming task well beyond the fast-mapping stage. It is also important to note that there was only one specific naming condition (out of six) that clearly exposed a difference between groups: the stationary condition of the location

game. It is not immediately clear why there would be an advantage for monolinguals in this condition. For the bilingual group, the mean for the variable condition—which should be more difficult—is actually higher than the mean for the stationary condition. This may suggest that the bilingual children were spending more cognitive resources on the variable condition, at the expense of the stationary condition, with naming being the weakest link in the word learning chain. Perhaps a rehearsal process was interrupted by focus on movement. These speculations require further study. Functionally, the fact that the difference clearly emerged on only one of six conditions suggests that, by the time that primarily simultaneous bilingual children are 7 years old, they have had enough experience with English phonology to be able to produce newly learned novel English-like nonwords, as well as their monolingual peers in most cases, although researchers may be able to identify vulnerable situations (e.g., Erikson et al., 2018).

We saw the largest monolingual advantage on a phonological task: mispronunciation detection. The bilingual children were less accurate than their monolingual peers on half of the games and had a tendency for five of the six games to be less accurate correctly rejecting foils. The foils that they heard and accepted in this study were phonologically related to the learned word, so they required a fairly detailed phonological representation to make an accurate choice. The foils only differed from the correct name by a manipulation of the final consonant (e.g., nudwef/nudweg/nudwev). Our interpretation is that this difference in accuracy is related to the fact that most bilingual individuals regularly hear different pronunciations of the same words in their environment and thus are more accepting of differences. For example, they may hear their English-speaking friends ask for a /kes[^]dIIa/ whereas their Spanish-speaking friends request a /kesadIja/. Bilingual

Table 10. Summary of ANCOVA results.

Ability Tested	Task	Learning Contexts					
		Phonology		VisuoSpatial		Orthography	Verbs
		Word Length	Phonological Similarity	Visual Similarity	Location		
Link phonological and semantic representations	Phonological-Visual Linking	ANECDOTAL (Favors no difference)	ANECDOTAL (Favors no difference) Follow-ups: MODERATE evidence for NO difference in both conditions	ANECDOTAL (Favors no difference)	MODERATE evidence for NO difference	ANECDOTAL (favors no difference)	MODERATE evidence for NO difference
Create and store a detailed phonological representation and recall and produce that representation	Naming	MODERATE evidence for NO difference	ANECDOTAL (Favors group difference) Follow-ups: ANECDOTAL (Monolingual advantage for similar); MODERATE evidence for NO difference for dissimilar	MODERATE evidence for NO difference	ANECDOTAL (Favors group difference) Follow-ups: MODERATE (Monolingual advantage for stationary); (NO evidence for difference for variable)	MODERATE evidence for NO difference	MODERATE evidence for NO difference
Create and store a detailed semantic representation and recall and produce that representation	Mispronunciation Detection	ANECDOTAL (Favors no difference)	ANECDOTAL (Favors no difference)	STRONG (Monolingual advantage)	STRONG (Monolingual advantage)	STRONG (Monolingual advantage)	ANECDOTAL (Favors no difference)
	Visual Difference Decision	MODERATE evidence for NO difference	MODERATE evidence for NO difference	MODERATE evidence for NO difference	ANECDOTAL (Favors bilingual group)	ANECDOTAL (Favors group difference) Follow-ups: VERY STRONG (Monolingual advantage for orthography present); MODERATE (evidence for NO difference for orthography absent)	MODERATE evidence for NO difference
	Visual Feature Recall	ANECDOTAL (Favors no difference)	MODERATE evidence for NO difference	MODERATE evidence for NO difference	ANECDOTAL (Favors bilingual group)	MODERATE evidence for NO difference	ANECDOTAL (Favors bilingual group)

children know that both refer to a tortilla filled with cheese, and the pronunciation difference does not change meaning to them. Sebastian-Galles (2010) makes this same case to account for the fact that other studies have found “reduced sensitivities to...word mispronunciations” in bilingual individuals (p. 253). She points to the presence of foreign-accented productions in the language environment of many bilingual individuals as a possible explanation for this “reduced sensitivity.” Accuracy on this task may be reduced compared to monolingual peers who hear relatively stable input. That stable input the monolinguals hear may lead to stricter boundaries between word pronunciations and productions. However, for the bilingual group, the variability that they hear may result in reduced sensitivity to these boundaries, allowing for broader variation in mapping the sounds that would correspond for any given word. This may be particularly true in situations where there are fewer resources available. It may not be coincidental that these differences only emerged in the situations that were more visually demanding, thus requiring more attention to the visual components of the words.

In either case, this monolingual “advantage” is likely not a functional advantage. That is, the bilingual children were able to accurately produce the names of the nonwords and link them to referents just as well as their monolingual peers in the majority of cases (11 of 12 instances). This would be impossible without a strong representation of the phonology of the labels. Their performance on the mispronunciation detection measure likely reflects the difference in their linguistic environment and thus their interpretation of the task. This is not a disadvantage, *per se*, but a difference. Researchers using discrimination tasks with bilingual participants should keep this in mind.

This monolingual advantage in mispronunciation detection was not expected based on the findings from Alt et al. (2013), in which there was no group difference between the monolingual and bilingual children. Why, then, was there no difference in discrimination performance for the children in the study of Alt et al. (2013)? One reason might be the differences in study samples. Alt et al. (2013) were interested in examining the effects of phonotactic probability, which is something children derive from input. Thus, their sample of bilingual children included children who heard English and Spanish. However, not all of the children in that sample were able to or chose to speak Spanish. Having more children in the sample who were more English-dominant might have made the group perform more like monolingual children. In that study, higher exposure to English did predict better performance on the task. In contrast, the children in the current study were intentionally selected to be functional bilinguals. That is, these children were overwhelmingly simultaneous bilinguals who maintained enough of their Spanish to be able to converse in Spanish in a region where there is a risk of attrition for Spanish. So, the difference may have been in the degree of bilingualism in each of the studies’ populations.

Although the literature did not provide strong predictions for what to expect for semantic tasks, it did suggest that monolingual and bilingual children might attend to different kinds of cues. We did not find clear evidence to support this hypothesis in general. The two potential advantages for bilingual children (visual feature recall in the location and verb games) were supported only by anecdotal evidence. The one place where there was an unambiguous monolingual advantage, though, was in the visual difference decision task for the orthography present condition. We interpreted Jubenville et al.’s (2014) work on orthography in bilingual children as showing that children were attending to different cues. Our findings seem to suggest that bilingual children might have more difficulty in processing semantic information if orthographic information is present. Clearly, there is no overall problem with processing semantic information, as seen by equivalent performance across the groups in the majority of other semantic conditions. It may be that orthography’s linguistic (i.e., graphemic and phonological) content interferes with learning the semantic information and it is learned when orthography is not present. The interference likely has something to do with the orthography itself, as the bilingual group’s performance on the naming and phonological–visual linking tasks show that they are able to learn phonological information in the context of semantic information easily.

Limitations

One limitation of this study is that we were not able to examine how bilingual children would perform on novel words with Spanish phonemes. We chose to focus on English phonemes to more closely model typical academic scenarios in the United States. However, this does not allow us to explore the direction of advantage/disadvantages or differences in bilingual word learning.

It might be cleaner to examine groups that shared more characteristics, but this may not always be possible within this population. For example, our groups differed in terms of SES. However, to use just one comparison as an example, in the United States, the poverty rate for Hispanics is 18.3% compared to 8.7% for non-Hispanic Whites (Federal Safety Net, 2018). Although it is certainly possible to match participants on SES, doing so, in some situations, might compromise the generalizability of the outcomes by selecting children who tend to be the outliers of their group in order to match the comparison group. We selected children from the same communities to eliminate as many extraneous differences as possible. We were surprised to find group difference in nonverbal intelligence scores, although please recall that both groups were well within normal limits. It is not clear what the cause of this difference was—our bilingual participants all understood English well enough to pass the training items on the assessment, and no language was used during testing. However, future work might try to either match participants on nonverbal intelligence, replicate these findings, or further

explore the reason for this difference. As we see in Appendix B, nonverbal intelligence clearly had an influence on the outcomes, although through the use of Bayesian statistics we were able to quantify the role of group independent of nonverbal intelligence.

Given the limited number of studies on the topic of novel word learning in bilingual school-age children, there is a near limitless number of possibilities to expand this line of work. First, we had anecdotal evidence on a full third of the tasks (10 or 30), meaning that it is difficult to determine if there is likely a group difference or not and that we cannot fully understand how learning is similar or different in these specific learning contexts. Another area for expansion is in the type of learners studied. For example, all the children in our study were proficient enough as users of their languages to hold a conversation in both languages. They all also had a similar age of acquisition for their language, given that over 90% of the children had exposure to both languages in their home. The patterns that we found might not hold for children with distinctly different levels of proficiency, later ages of acquisition, or other differences in their bilingualism.

Finally, given the limited literature on this topic, our study was necessarily somewhat exploratory. This is particularly the case for the semantic component of the tasks. Our comprehensive design allowed us to examine word learning using five different tasks in six conditions, allowing us to compare across 30 unique situations. However, we had only a partial set of expectations, making the study partly confirmatory and partly exploratory.

Conclusion

This study is one of only a few examinations of novel word learning in monolingual and bilingual school-age children. Our study allowed us to look at word learning in a more comprehensive manner than previous studies of word learning by examining performance on five different word learning tasks across six different games that had different learning challenges. We extend the findings from previous work with results from tasks tapping into phonological linking, phonological knowledge, semantic knowledge, and the way this knowledge is applied across six different word learning conditions. There were no group differences in half of our tasks, but we uncovered some important differences in the groups' learning. Some of these differences clearly warrant further exploration. For example, bilingual children's more inclusive interpretation of differences in pronunciations could have direct consequences in an English-based academic setting. Specifically, instructors need to be aware of the potential for bilingual children to be more open to alternate pronunciations and should be clear, especially when teaching novel words with similar pronunciations (e.g., adductor/abductor), about which phonemic differences are meaningful. The relative difficulty learning semantic content when orthography was present is worth further study. Before educators rush to eliminate orthography from instructional material

in fear of impeding content knowledge, we should remember that, even though this was the less advantageous condition for the bilingual group, the score in that condition was not substantially different than the scores for other games. Thus, this "disadvantage" is relative and should be interpreted with caution. The bottom line is that the groups, despite differences in SES and nonverbal IQ, had similar performance on the majority of measures. Thus, if a bilingual child who has a profile similar to the children in our study exhibits a generalized difficulty with word learning, it would suggest a true learning problem and not something to be expected due to bilingualism. There clearly is far more work to be done on bilingual novel word learning to fully understand the mechanisms involved.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R01 DC010784, awarded to Shelley Gray. We are deeply grateful to the staff, research associates, school administrators, teachers, children, and families who participated. Key personnel included (in alphabetical order) Shara Brinkley, Gary Carstensen, Cecilia Figueroa, Karen Guilmette, Trudy Kuo, Bjorg LeSueur, Annelise Pesch, and Jean Zimmer. Many students also contributed to this work including (in alphabetical order) Lauren Baron, Alexander Brown, Nora Schlesinger, Nisha Talanki, and Hui-Chun Yang.

References

- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*, 207–245. <https://doi.org/10.3102/0034654310368803>
- Alt, M., Arizmendi, G. D., & DiLallo, J. N. (2016). The role of socioeconomic status in the narrative story retells of school-aged English language learners. *Language, Speech, and Hearing Services in Schools, 47*, 313–323. https://doi.org/10.1044/2016_LSHSS-15-0036
- Alt, M., Hogan, T., Green, S., Gray, S., Cabbage, K., & Cowan, N. (2017). Word learning deficits in children with dyslexia. *Journal of Speech, Language, and Hearing Research, 60*, 1012–1028. https://doi.org/10.1044/2016_JSLHR-L-16-0036
- Alt, M., Meyers, C., & Figueroa, C. (2013). Factors that influence fast-mapping in children exposed to Spanish and English. *Journal of Speech, Language, and Hearing Research, 56*, 1237–1248. [https://doi.org/10.1044/1092-4388\(2012/11-0092\)](https://doi.org/10.1044/1092-4388(2012/11-0092))
- Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research, 47*, 407–420. [https://doi.org/10.1044/1092-4388\(2004/033\)](https://doi.org/10.1044/1092-4388(2004/033))
- Arizmendi, G. D., Alt, M., Gray, S., Hogan, T. P., Green, S., & Cowan, N. (2018). Do bilingual children have an executive function advantage? Results from inhibition, shifting, and updating tasks. *Language, Speech, and Hearing Services in Schools, 49*, 356–378. https://doi.org/10.1044/2018_LSHSS-17-0107
- Baron, L. S., Hogan, T. P., Alt, M., Gray, S., Cabbage, K. L., Green, S., & Cowan, N. (2018). Children with dyslexia benefit from orthographic facilitation during spoken word learning. *Journal of Speech, Language, and Hearing Research, 61*, 2002–2014. https://doi.org/10.1044/2018_JSLHR-L-17-0336

- Barragan, B., Castilla-Earls, A., Martinez-Nieto, L., Restrepo, M. A., & Gray, S. (2018). Performance of low-income dual language learners attending English-only schools on the Clinical Evaluation of Language Fundamentals–Fourth Edition, Spanish. *Language, Speech, and Hearing Services in Schools*, 48, 292–305. https://doi.org/10.1044/2017_LSHSS-17-0013
- Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition*, 15, 397–401. <https://doi.org/10.1017/S136672891100040X>
- Blumenfeld, H. K., & Marian, V. (2009). Language–cognition interactions during bilingual language development in children. In B. Kuzmanovic & A. Cuevas (Eds.) *Recent trends in education* (pp. 39–69). Hauppauge, NY: Nova Science.
- Brojde, C. L., Ahmed, S., & Colunga, E. (2012). Bilingual and monolingual children attend to different cues when learning new words. *Frontiers in Psychology*, 3, 155. <https://doi.org/10.3389/fpsyg.2012.00155>
- Brownell, R. (2001). *Expressive One-Word Picture Vocabulary Test–4: Spanish-Bilingual Edition (EOWPVT-4: SBE)*. Novato, CA: Academic Therapy Publications.
- Buac, M., Gross, M., & Kaushanskaya, M. (2016). Predictors of processing-based task performance in bilingual and monolingual children. *Journal of Communication Disorders*, 62, 12–29. <https://doi.org/10.1016/j.jcomdis.2016.04.001>
- Byers-Heinlein, K., Fennell, C. T., & Werker, J. F. (2013). The development of associative word learning in monolingual and bilingual infants. *Bilingualism: Language and Cognition*, 16, 198–205. <https://doi.org/10.1017/S1366728912000417>
- Cabbage, K. L., Brinkley, S., Gray, S., Alt, M., Cowan, N., Green, S., . . . Hogan, T. (2017). Assessing working memory in children: The Comprehensive Assessment Battery for Children–Working Memory (CABC-WM). *Journal of Visualized Experiments*, 124, e55121. <https://doi.org/10.3791/55121>
- Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, 130, 278–288. <https://doi.org/10.1016/j.cognition.2013.11.015>
- Carlisle, J. F., Fleming, J. E., & Gudbrandsen, B. (2000). Incidental word learning in science classes. *Contemporary Educational Psychology*, 25, 184–211. <https://doi.org/10.1006/ceps.1998.1001>
- Cowan, N., Hogan, T. P., Alt, M., Green, S., Cabbage, K. L., Brinkley, S., & Gray, S. (2017). Short-term memory in childhood dyslexia: Deficient serial order in multiple modalities. *Dyslexia*, 23, 209–233. <https://doi.org/10.1002/dys.1557>
- De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children’s vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, 35, 1189–1211.
- DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M. J., & McGoey, K. E. (1998). Parent ratings of attention-deficit/hyperactivity disorder symptoms: Factor structure and normative data. *Journal of Psychopathology and Behavioral Assessment*, 20, 83–102.
- Erikson, J. A., Alt, M., Gray, S., Green, S., Hogan, T. P., & Cowan, N. (2018). Phonological vulnerability for school-aged Spanish–English speaking bilingual children. *International Journal of Bilingual Education and Bilingualism*. Retrieved from <http://www.tandfonline.com/doi/full/10.1080/13670050.2018.1510892>
- Etz, A., Gronau, Q. F., Dablander, F., Edelsbrunner, P. A., & Baribault, B. (2018). How to become a Bayesian in eight easy steps: An annotated reading list. *Psychonomic Bulletin & Review*, 25, 219–234. <https://doi.org/10.3758/s13423-017-1317-5>
- Federal Safety Net. (2018). *U.S. poverty statistics*. Retrieved from <http://federal-safetynet.com/us-poverty-statistics.html>
- Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2)*. Circle Pines, MN: AGS.
- Goldstein, B. A. (2004). *Bilingual language development and disorders in Spanish–English speakers*. Baltimore, MD: Brookes.
- Gray, S., Green, S., Alt, M., Hogan, T., Kuo, T., Brinkley, S., & Cowan, N. (2017). The structure of working memory in young school-age children and its relation to intelligence. *Journal of Memory and Language*, 92, 183–201. <https://doi.org/10.1016/j.jml.2016.06.004>
- Green, S. B., Yang, Y., Alt, M., Brinkley, S., Gray, S., Hogan, T., & Cowan, N. (2016). Use of internal consistency coefficients for estimating reliability of experimental task scores. *Psychonomic Bulletin and Review*, 23, 750–763. <https://doi.org/10.3758/s13423-015-0968-3>
- Han, W. J. (2012). Bilingualism and academic achievement. *Child Development*, 83, 300–321. <https://doi.org/10.1111/j.1467-8624.2011.01686.x>
- Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, 49, 4–14. <https://doi.org/10.1037/a0027238>
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39, 1–27. <https://doi.org/10.1017/S0305000910000759>
- Houston-Price, C., Caloghris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, 15, 125–150. <https://doi.org/10.1111/j.1532-7078.2009.00009.x>
- JASP Team. (2018). JASP (Version 0.8.6) [Computer software]. Retrieved from <https://jasp-stats.org/2018/02/08/now-jasp-0-8-6/>
- Jubenville, K., Sénéchal, M., & Malette, M. (2014). The moderating effect of orthographic consistency on oral vocabulary learning in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 126, 245–263. <https://doi.org/10.1016/j.jecp.2014.05.002>
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition (K-ABC2)*. Circle Pines, MN: AGS.
- Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, 16, 705–710. <https://doi.org/10.3758/PBR.16.4.705>
- Kruschke, J. K. (2013). Bayesian estimation supersedes the *t* test. *Journal of Experimental Psychology: General*, 142, 573–603. <https://doi.org/10.1037/a0029146>
- Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, 55, 306–353. <https://doi.org/10.1016/j.cogpsych.2007.01.001>
- Mattock, K., Polka, L., Ryachew, S., & Krehm, M. (2010). The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science*, 13, 229–243. <https://doi.org/10.1111/j.1467-7687.2009.00891.x>
- McGregor, K. K., Friedman, R. M., Reilly, R. M., & Newman, R. M. (2002). Semantic representation and naming in young children. *Journal of Speech, Language, and Hearing Research*, 45, 332–346. [https://doi.org/10.1044/1092-4388\(2002\)026](https://doi.org/10.1044/1092-4388(2002)026)
- Merriman, W. E., & Kutlesic, V. (1993). Bilingual and monolingual children’s use of two lexical acquisition heuristics.

- Applied Psycholinguistics*, 14(2), 229–249. <https://doi.org/10.1017/S0142716400009565>
- Morales, J., Calvo, A., & Bialystok, E.** (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114, 187–202. <https://doi.org/10.1016/j.jecp.2012.09.002>
- Nair, vs. K. K., Biedermann, B., & Nickels, L.** (2016). Consequences of late bilingualism for novel word learning: Evidence from Tamil–English bilingual speakers. *International Journal of Bilingualism*, 20, 473–487. <https://doi.org/10.1177/1367006914567005>
- Ouellette, G. P.** (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98, 554–566. <https://doi.org/10.1037/0022-0663.98.3.554>
- Paradis, J., & Genesee, F.** (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition*, 18, 1–25. <https://doi.org/10.1017/S0272263100014662>
- Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M.** (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary. *Reading & Writing Quarterly*, 31, 235–252. <https://doi.org/10.1080/10573569.2015.1030995>
- Ricketts, J., Bishop, D. v. M., & Nation, K.** (2009). Orthographic facilitation in oral vocabulary acquisition. *The Quarterly Journal of Experimental Psychology*, 62, 1948–1966. <https://doi.org/10.1080/17470210802696104>
- Rojas, R., & Iglesias, A.** (2013). The language growth of Spanish-speaking English language learners. *Child Development*, 84, 630–646. <https://doi.org/10.1111/j.1467-8624.2012.01871.x>
- Sebastian-Galles, N.** (2010). Bilingual language acquisition: Where does the difference lie? *Human Development*, 53, 245–255. <https://doi.org/10.1159/000321282>
- Semel, E., Wiig, E. H., & Secord, W. A.** (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4)*. Circle Pines, MN: AGS.
- Semel, E., Wiig, E. H., & Secord, W. A.** (2006). *Clinical Evaluation of Language Fundamentals–Fourth Edition, Spanish (CELF-4 Spanish)*. San Antonio, TX: Pearson.
- Tabri, D., Abou Chacra, K. M. S., & Pring, T.** (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International Journal of Language & Communication Disorders*, 46, 411–422.
- The Nation's Report Card.** (2015). *Mathematics and reading assessments*. Retrieved from http://www.nationsreportcard.gov/reading_math_2015/#reading/groups?grade=4
- Torgesen, J., Wagner, R., & Rashotte, C.** (2012). *Test of Word Reading Efficiency–Second Edition (TOWRE-2)*. Circle Pines, MN: AGS.
- Vitevitch, M. S., & Luce, P. A.** (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36, 481–487.
- Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D.** (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25, 58–76.
- Williams, K. T.** (2007). *Expressive Vocabulary Test–Second Edition*. San Antonio, TX: Pearson.
- Woodcock, R.** (2011). *Woodcock Reading Mastery Test–Third Edition*. San Antonio, TX: Pearson.

Appendix A

Nonword Stimuli Characteristics

Condition of interest	Nonword in Klattese ^a	Condition manipulation	Duration in milliseconds with means and SDs in gray ^b	Biphone frequency with means and SDs in gray ^c	Summed biphone probability with means and SDs in gray ^d	Similar foils ^e	Dissimilar foils ^f
Phonological similarity	<i>n^dwef</i> <i>w^gyed</i>	Similar Similar	845 825	0.0008 0.0009	0.0039 0.0043	<i>nudwev</i> <i>w^gyet</i>	<i>nudweg</i> <i>w^gyef</i>
			835.00 (14.14)	0.0009 (0.0001)	0.0041 (0.0002)		
	<i>hWktcf</i> <i>gomp@v</i>	Dissimilar Dissimilar	1111 873	0.0012 0.0028	0.0050 0.0140	<i>hWktcv</i> <i>gomp@f</i>	<i>hWktcg</i> <i>gomp@p</i>
<i>M (SD)</i>			992.00 (168.29)	0.0020 (0.0011)	0.0095 (0.0063)		
Word length	<i>kYmtUp</i> <i>dUdtif</i>	Short Short	867 850	0.0004 0.0012	0.0022 0.0061	<i>kYmtUb</i> <i>dUdtiv</i>	<i>kYmtUz</i> <i>dUdtig</i>
			858.50 (12.02)	0.0008 (0.0006)	0.0041 (0.0027)		
	<i>wefyUktughcd</i> <i>nUdfegdYnyup</i>	Long Long	1600 1585	0.0015 0.0004	0.0162 0.0044	<i>wefyUktughcn</i> <i>nUdfegdYnyub</i>	<i>wefyUktughcf</i> <i>nUdfegdYnyun</i>
<i>M (SD)</i>			1592.50 (10.61)	0.0010 (0.0008)	0.0103 (0.0083)		
Visual similarity	<i>dofwig</i> <i>b^vdep</i> <i>yitgYm</i> <i>fugbOn</i>	†	875 736 772 917	0.0004 0.0012 0.0007 0.0003	0.0022 0.0059 0.0034 0.0015	<i>dofwik</i> <i>b^vdeb</i> <i>yitgYn</i> <i>fugbOd</i>	<i>dofwim</i> <i>b^vden</i> <i>yitgYk</i> <i>fugbOk</i>
<i>M (SD)</i>			825.00 (85.04)	0.00065 (0.0004)	0.0032 (0.0019)		
Grid location	<i>t^pwib</i> <i>tughWt</i> <i>kYmyeg</i> <i>yiktuf</i>	†	958 783 904 999	0.0006 0.0006 0.0006 0.0015	0.0032 0.0028 0.0028 0.0077	<i>t^pwim</i> <i>tughWd</i> <i>kYmyek</i> <i>yiktuv</i>	<i>t^pwin</i> <i>tughWv</i> <i>kYmyen</i> <i>yiktug</i>
<i>M (SD)</i>			911.00 (93.78)	0.00083 (0.0004)	1.0041 (0.0023)		
Orthography	<i>banfep</i> <i>m^bglk</i> <i>dimbYg</i> <i>duftcg</i>	†	830 1057 931 902	0.0038 0.0041 0.0018 0.0009	0.0188 0.0204 0.0090 0.0044	<i>banfeb</i> <i>M^bglg</i> <i>dimbYk</i> <i>duftck</i>	<i>banfen</i> <i>M^bglf</i> <i>dimbYm</i> <i>duftcf</i>
<i>M (SD)</i>			930.00 (94.72)	0.0026 (0.0016)	0.0131 (0.0077)		
Verbs	<i>buvyid</i> <i>hOtgyM</i> <i>kibtUp</i> <i>yevhWt</i>	†	900 939 840 915	0.0004 0.0002 0.0003 0.0004	0.0020 0.0010 0.0017 0.0020	<i>buvyip</i> <i>hOtgyN</i> <i>kibtUb</i> <i>yevhWd</i>	<i>buvyik</i> <i>hOtgyk</i> <i>kibtUz</i> <i>yevhWv</i>
<i>M (SD)</i>			898.50 (42.18)	0.0003 (0.0001)	0.0017 (0.0005)		

Note. † indicates the computer randomly assigned two nonword–monster pairs to each condition for all participants.

^aKlattese is a computer-readable interface for the International Phonetic Alphabet. See Vitevitch and Luce (2004) for more information.

^bExcluding the “long” phonological manipulation, there was no significant effect of duration on condition of interest when using multiple *t* tests for independent samples and a Bonferroni correction for multiple comparisons ($p < .005$). ^cAll phonotactic probabilities were calculated using the phonotactic probability calculator (Vitevitch & Luce, 2004). There was no difference between biphone frequency means for any condition when using multiple *t* tests for independent samples and a Bonferroni correction for multiple comparisons ($p < .003$). ^dThere was no difference between summed biphone probabilities for any condition when using multiple *t* tests for independent samples and a Bonferroni correction for multiple comparisons ($p < .005$). Long words were not included in this comparison because they have more phonemes and thus higher biphone probabilities. ^eSimilar foils differed from the target word by a single consonant feature in the word-final phoneme (12 differed in voicing, three differed in manner, one differed in place). Foils were only presented during the mispronunciation decision task. ^fDissimilar foils primarily differed from the target word by all three consonant features in the word-final phoneme (three differed by two consonant features only). Foils were only presented during the mispronunciation decision task.

Appendix B (p. 1 of 7)

Bayes Inclusion Factors for All Primary Effects for All Tasks

The following are analyses of effects that compares effects across matched models. That is, these findings compare “models that contain the effect to equivalent models stripped of the effect. Higher order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.” (JASP 0.8.6; JASP Team, 2018). KABC = Kaufman Assessment Battery for Children.

B1. Phonological–Visual Linking Task

Word Length

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Length	0.400	0.951	2.192e+6
Group	0.400	0.250	0.357
KABC	0.500	0.933	14.014
Length × Group	0.200	0.049	0.197

Phonological Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Similarity	0.400	0.117	0.146
Group	0.400	0.311	0.511
KABC	0.500	0.991	104.279
Similarity × Group	0.200	0.080	2.149

Visual Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Similarity	0.400	0.925	26.875
Group	0.400	0.258	0.368
KABC	0.500	0.998	510.905
Similarity × Group	0.200	0.041	0.165

Location

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Location	0.400	0.113	0.128
Group	0.400	0.236	0.311
KABC	0.500	0.965	27.551
Location × Group	0.200	0.004	0.166

Appendix B (p. 2 of 7)

Bayes Inclusion Factors for All Primary Effects for All Tasks

Orthography

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Orthography	0.400	0.918	559.452
KABC	0.500	0.986	71.997
Group	0.400	0.455	0.980
Orthography × Group	0.200	0.080	0.177

Verbs

Analysis of Effects - LR Mean

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Group	0.500	0.158	0.187
KABC	0.500	0.898	8.839

B2. Naming Task

Word Length

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Length	0.400	0.929	5.725e+41
Group	0.400	0.149	0.192
KABC	0.500	0.981	52.348
Length × Group	0.200	0.071	0.472

Phonological Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Similarity	0.400	0.092	0.104
Group	0.400	0.248	0.341
KABC	0.500	0.681	2.138
Similarity × Group	0.200	0.025	1.079

Visual Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Visual Similarity	0.400	0.971	278.494
Group	0.400	0.134	0.159
KABC	0.500	0.992	125.232
Visual Similarity × Group	0.200	0.026	0.192

Appendix B (p. 3 of 7)

Bayes Inclusion Factors for All Primary Effects for All Tasks

Location

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Location	0.400	0.247	0.383
Group	0.400	0.313	0.542
KABC	0.500	0.826	4.759
Location × Group	0.200	0.110	1.168

Orthography

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Orthography	0.400	0.974	4.995e+19
Group	0.400	0.159	0.194
KABC	0.500	0.996	240.429
Orthography × Group	0.200	0.026	0.161

Verbs

Analysis of Effects—Naming Mean

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Group	0.500	0.243	0.321
KABC	0.500	0.993	146.427

B3. Mispronunciation Detection Task

Word Length

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Group	0.400	0.243	0.343
KABC	0.500	0.933	14.033
Length	0.400	0.950	2.134e+6
Group × Length	0.200	0.050	0.207

Phonological Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Phono Sim	0.400	0.130	0.151
Group	0.400	0.394	0.659
KABC	0.500	0.931	13.441
Phono Sim × Group	0.200	0.009	0.170

Appendix B (p. 4 of 7)

Bayes Inclusion Factors for All Primary Effects for All Tasks

Visual Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Visual Similarity	0.400	0.186	0.253
Group	0.400	0.873	18.536
KABC	0.500	0.936	14.587
Visual Similarity × Group	0.200	0.080	0.453

Location

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Location	0.400	0.147	0.179
Group	0.400	0.883	10.366
KABC	0.500	0.572	1.334
Location × Group	0.200	0.032	0.235

Orthography

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Orthography	0.400	0.209	0.278
Group	0.400	0.891	12.555
KABC	0.500	0.337	0.508
Orthography × Group	0.200	0.038	0.194

Verbs

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Group	0.500	0.293	0.414
KABC	0.500	0.870	6.707

B4. Visual Difference Decision Task

Word Length

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Length	0.400	0.094	0.105
Group	0.400	0.201	0.254
KABC	0.500	0.979	46.460
Length × Group	0.200	0.007	0.351

Appendix B (p. 5 of 7)

Bayes Inclusion Factors for All Primary Effects for All Tasks

Phonological Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Phono Similarity	0.400	0.127	0.147
Group	0.400	0.212	0.275
KABC	0.500	0.598	1.485
Phono Similarity × Group	0.200	0.014	0.542

Visual Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Visual Similarity	0.400	0.977	100049.374
Group	0.400	0.138	0.165
KABC	0.500	0.698	2.315
Visual Similarity × Group	0.200	0.023	0.165

Location

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Location	0.400	0.196	0.250
Group	0.400	0.550	1.274
KABC	0.500	0.947	17.841
Location × Group	0.200	0.019	0.174

Orthography

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Orthography	0.400	0.108	0.133
Group	0.400	0.272	0.418
KABC	0.500	0.995	184.998
Orthography × Group	0.200	0.077	2.293

Verbs

Analysis of Effects—SEM

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Group	0.500	0.159	0.189
KABC	0.500	0.998	539.033

Appendix B (p. 6 of 7)

Bayes Inclusion Factors for All Primary Effects for All Tasks

B5. Visual Feature Recall Task

Word Length

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Length	0.400	0.104	0.117
Group	0.400	0.254	0.344
KABC	0.500	1.000	7639.765
Length × Group	0.200	0.005	0.198

Phonological Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Similarity	0.400	0.122	0.140
Group	0.400	0.251	0.337
KABC	0.500	1.000	2539.004
Similarity × Group	0.200	0.006	0.244

Visual Similarity

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Similarity	0.400	0.943	67.703
Group	0.400	0.147	0.181
KABC	0.500	0.998	548.170
Similarity × Group	0.200	0.043	0.297

Location

Analysis of Effects

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Location	0.400	0.118	0.135
Group	0.400	0.511	1.064
KABC	0.500	0.998	626.979
Location × Group	0.200	0.009	0.151

Appendix B (p. 7 of 7)**Bayes Inclusion Factors for All Primary Effects for All Tasks**

Orthography**Analysis of Effects**

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Orthography	0.400	0.099	0.110
Group	0.400	0.168	0.203
KABC	0.500	0.856	5.954
Orthography × Group	0.200	0.003	0.175

Verbs**Analysis of Effects—VFR Mean**

Effects	P(incl)	P(inclldata)	BF _{Inclusion}
Group	0.500	0.567	1.312
KABC	0.500	0.604	1.523