Published in final edited form as:
Sci Stud Read. 2019 ; 23(1): 49–63. doi:10.1080/10888438.2018.1466303.

Development and Prediction of Context-Dependent Vowel Pronunciation in Elementary Readers

Laura M. Steacy,
Florida Center for Reading Research, Florida State University

Donald L. Compton,
Florida Center for Reading Research, Florida State University

Yaacov Petscher,
Florida Center for Reading Research, Florida State University

James D. Elliott,
Florida Center for Reading Research, Florida State University

Kathryn Smith,
Florida Center for Reading Research, Florida State University

Jay G. Rueckl,
University of Connecticut & Haskins Laboratories

Oliver Sawi,
University of Connecticut & Haskins Laboratories

Stephen J. Frost, and
Haskins Laboratories

Kenneth R. Pugh
Haskins Laboratories

Abstract

As children learn to read they become sensitive to context-dependent vowel pronunciations in words, considered a form of statistical learning. The work of Treiman and colleagues demonstrated that readers' vowel pronunciations depend on the consonantal context in which the vowel occurs and reading experience. We examined child- and nonword-factors associated with children's assignment of more vs. less frequent grapheme-phoneme correspondences (GPC) to vowel pronunciations as a function of rime coda in monosyllabic nonwords. Students (N=96) in grades 2–5 read nonwords in which more vs. less frequent vowel GPCs were wholly supported or partially favored by the rime unit. Two explanatory item-response models were developed using alternative nonword scoring procedures. Use of less frequent vowel GPCs was predicted by set for variability, word reading, and rime support for the context-dependent vowel pronunciation. We interpret the results within a developmental word reading model in which initially incomplete and oversimplified GPC representations become more context-dependent with reading experience.

Correspondence concerning this article should be sent to Laura M. Steacy: lsteacy@fcrr.org.

Acquiring a system of lexical representations that permit efficient word recognition is an essential part of learning to read in any language (see Daniels & Share, in press; Ehri, 2014; Perfetti & Stafura, 2014). The transparency of the mapping system relating orthography to phonology (O-P) in alphabetic scripts affects the ease with which initial orthographic representations are formed in developing readers. For instance, Seymour, Aro, and Erskine (2003) reported that children who were acquiring reading in languages with transparent mappings between printed and spoken forms (e.g., Greek, Finnish, German, Italian, Spanish) were close to ceiling in monosyllable word reading by the middle of first grade, whereas English-speaking children performed extremely poorly (34% correct). Danish (71%), Portuguese (73%), and French (79%) children showed somewhat lower levels of word reading accuracy, which is in line with the reduced transparency of these languages.

English is an example of a *quasi-regular* orthography because the mapping between written and spoken words contains substantial ambiguity (see Kessler, 2009; Plaut, McClelland, Seidenberg, & Patterson, 1996). In contrast to a transparent orthography (e.g., Spanish), where there is a nearly one-to-one mapping between letters and phonemes, in English, phonemes can be represented by either single letters or letter cluster (e.g., *-ph* in *graph*), and many graphemes can be pronounced in more than one way (cf. *pint* vs *hint, bead* vs *head*). The Seymour et al. (2003) study suggests that the inconsistency in the mapping of O-P poses significant challenges to the beginning readers of English. Readers of transparent orthographies can learn to decode by acquiring a set of representations linking each grapheme with a particular phoneme and then applying these grapheme-phoneme correspondences (GPC) in a left-to-right fashion, thus assembling the pronunciation of a written word (see Coltheart et al., 2001). Quasi-regular orthographies such as English are problematic for a developing reader relying on rules of this sort, because not all words can be correctly pronounced using GPC rules of the type just described (e.g., words such as *head* and *pint*). One possibility is that readers rely on a whole-word recognition process (i.e., retrieved ‘directly’ via the lexical pathway), rather than decoding, to read exception words. While there is substantial evidence for this account (Coltheart et al., 2001), including data that readers of less consistent orthographies rely more on the lexical pathway (Frost et al., 1987; Ziegler et al., 2010), there is also evolving evidence for another solution to the challenges imposed by quasi-regular writing systems—the use of O-P correspondences that are sensitive to context (Treiman, Kessler, & Bick, 2003; Treiman, Kessler, Zevin, Bick, & Davis, 2006; Ziegler & Goswami, 2005).

In English, much of the ambiguity associated with the pronunciation of a particular grapheme, in particular vowels, can be resolved by considering the context in which they occur (see Venezky, 1999). For instance, *ea* is pronounced as /i/ in *beat*, /ɛ/ in *head*, and /eɪ/ in *steak*. Because /i/ is the most frequent of these pronunciations, a decoding system that operates on each grapheme independently would misread *head* and *steak.* However, if the following consonant (i.e., the coda) is taken into account, then /i/ is the most frequent pronunciation in *–eat*, but /ɛ/ is the most frequent pronunciation in *–ead.* In a corpus analysis, Kessler and Treiman (2001) found that the consistency of vowel pronunciations increases significantly when the syllable coda is considered. Specifically, the spelling-sound consistency of vowels in monosyllabic words was .71 when considering the unconditioned vowel (i.e., no consideration of consonantal context), but .92 when the coda was used to condition the vowel (i.e., consideration of the consonantal context), a difference that is statistically significant. Thus, a decoding process based on multi-grapheme units could successfully decode both *beat* and *head*. (Note that *steak* would still be misread. In some cases, the only context that reliably indicates the correct pronunciation is the whole word.)

There is substantial evidence that both children and adult readers of English use knowledge of regularities involving units larger than individual graphemes and phonemes (Johnson & Venezky, 1976; Glushko, 1979; Ryder & Pearson, 1980; Andrews & Scarratt, 1988). This type of learning can be considered statistical, in the sense that it is based on implicit analysis: orthographic patterns are essentially observations (or computations) that particular O-P correspondences are more frequent than others in a particular context (Kessler, 2009). To test this statistical learning hypothesis, Treiman et al. (2003; 2006) asked developing readers and adults to pronounce nonwords in which the experimental context for each case was the one in which the critical vowel pronunciation occurs most often in real English words, but is the less frequent GPC for the vowel (e.g., *oo* following the final /*k*/ as in *book*). There was also a control context in which the vowel, in this example *oo,* is pronounced in the typical manner—before final consonants such as /*m*/ and /*n*/ (e.g., *room* and *noon*) – corresponding to the more frequent GPC for the vowel.

Treiman et al. (2003; 2006) observed that vowel pronunciation (e.g., *oo*) in a nonword depended on the context in which it occurs (e.g., *pook* vs. *poom*). Before most consonants, *oo* is pronounced as /u/. This pronunciation occurs in words such as *room*, *soon*, and *hoot*. A different pronunciation, /ʊ/, is more common before /k/, as in *book*, *cook*, and *hook*. Just a few words, such as *spook*, have the /u/ pronunciation of *oo* before /k/. Thus, whereas *poom* is almost always read as rhyming with *room*, *pook* is sometimes read as rhyming with *spook* and sometimes as rhyming with *book*, suggesting that the decoding process is sensitive to the context in which a grapheme appears. The Treiman et al. studies demonstrate that this sensitivity to grapheme context develops early (i.e., first grade), continues through elementary school (i.e., fifth grade), and is most pronounced in adults. In addition, children's use of consonantal context correlated significantly with standardized word reading performance across a large grade range, indicating a robust relationship between inductive learning of O-P statistical relationships and general word reading ability. Results of the Treiman et al. (2006) study support the conventional wisdom that O-P connections in developing readers are initially based on simple one-to-one correspondences that are relatively insensitive to orthographic context (see Share, 1995) and that these initially incomplete and oversimplified representations become more sophisticated context-dependent connections as a result of reading experience. While the Treiman et al. (2003) results indicate a gradual shift towards the use of context-dependent O-P connections, children's understanding of O-P relations evolve quickly with reading experience (i.e., first grade) to represent context-dependent connections.

Overall, the results of Treiman et al. (2006) suggest that children become sensitive to the statistical regularities representing context-dependent O-P relationships that exist in the English orthography. However, Kessler (2009) and Treiman et al. (2003; 2006) provide two important caveats regarding the results that helped motivate the current study. First, the rate of critical vowel pronunciations (i.e., rate of statistical pattern use) in experimental nonwords (e.g., *pook* pronounced as /pʊk/), in both children and adults, was lower than that found in the general corpus of words. Treiman et al. offer several possible explanations for this discrepancy between corpus statistics and participant sensitivity to contextual constraints. One is that even though the critical pronunciations dominate in certain contexts, they are in the minority overall. Treiman et al. (2003) hypothesize that while *chead* shares –*ead* with *dead* and *head*, promoting use of /ɛ/, it also shares units with *cheap* and *meat*, promoting /i/. As such, the rate of /ɛ/ pronunciations among real words with -*ead* may not be sufficient to totally counteract the competing pronunciation of /i/ in words that share –*eat* and –*eap*. Thus, use of the minority critical vowel pronunciation (i.e., *ea* as /ɛ/) versus the majority vowel pronunciation (i.e., *ea* as /i/) is likely the result of some form of trade-off between vowel GPC frequency and strength of context-dependent O-P relationships in the rime unit throughout the corpus of English words. Another possible explanation is that children are typically taught the high frequency pronunciation of the vowel in phonics lessons (i.e., *ea* as /i/) and are therefore biased towards using the most frequent GPC for vowels due to instruction.

The second caveat is that children at lower reading levels (in particular first grade readers) did not show universal effects across all of the eight nonword patterns in the experiments. At lower reading levels, in particular first grade readers, children did not show universal effects across nonword types. Specifically, they pronounced *o* as /o/ more often before –*ld* and –*lt* than before other coda letters in the absence of final *e*; and they pronounced *oo* as /ʊ/ more often before -*k* than before other coda letters. At higher reading levels, children were sensitive to all eight patterns in the experiments, although significant variability persisted across the categories. Similarly, in a spelling experiment using a comparable methodology with developing spellers (Treiman and Kessler, 2006), participants' first use of the conditional vowel pronunciation correlated with the frequency of the most common context independent spelling of the vowel in question, with context sensitive spelling emerging first in vowels with the least consistent pronunciations. Kessler (2009) speculates that young children are more likely to pay attention to context when there is no candidate that clearly dominates across contexts, suggesting that there is a "payoff" for learning to use conditional relationships when the vowel GPC is highly variant (p. 30).

Taken together the Treiman et al. (2003; 2006) findings seem to suggest that the pronunciation of variant vowels in nonwords by developing readers are driven by both generalizations of vowel GPC frequency (the most common pronunciation of a vowel across all contexts) and support of context specific O-P relationships (how vowels are pronounced in particular contexts) within the corpus of English words. However, little is known about how these two knowledge sources compete during the pronunciation of nonwords with and without context-dependent connections in developing readers. In the current study, we extend the Treiman et al. (2006) study by considering a more diverse set of nonwords, partitioning item variance across nonwords and participants using explanatory item-response models (a form of IRT), and including a diverse set of participant predictors (e.g. phonemic awareness, rapid automatized naming, set for variability, visual statistical learning, vocabulary, and reading skill) and a nonword predictor (a continuous measure of the support for the alternative pronunciation of each item based on a type ratio between words containing the rime with the conditional vowel pronunciation to the total occurrences of the rime). We were specifically interested in trying to understand why some developing readers may be more willing to consider context-dependent O-P relations when reading nonwords and why certain rime patterns may have a stronger influence on supporting context-dependent vowel pronunciations than others.

To accomplish this, we present two statistical models using different coding procedures, the first focusing on the use of unconditionalized higher frequency vowel GPC pronunciation (HF-GPC) of the nonwords and the second on the use of conditionalized lower frequency, rime influenced, vowel GPC pronunciation (LF-GPC) of the nonwords. (Two models were necessary because results from nonword pronunciations can be considered as correct based on the unconditionalized vowel pronunciation, correct based on the conditionalized vowel pronunciation, or incorrect). However, it is important to note that the use of the two coding schemes to capture competing vowel pronunciations in no way connotes different processes for reading nonwords. Rather, we adhere to the theoretical perspective that there is one process involved in the naming of all nonwords. From this perspective, nonword pronunciation is constrained by statistical regularities involving both grapheme-phoneme and rime correspondences, and the likelihood that rime correspondences regularities will win out over GPC regularities (in cases where they conflict) depends on both the strength of the rime correspondences in the corpus and child-specific factors. Contrasting results across the two analyses allows an exploration of the trade-offs in developing readers between generalizations based on vowel GPC frequency (the most common pronunciation of a vowel across all contexts) and context specific O-P relationships (how vowels are pronounced in particular rime contexts) when reading nonwords with variant vowels[1]. Predictors were chosen to help identify child- and nonword-attributes that explain why certain children are more prone to use alternative vowel pronunciations associated with the rime coda and why certain nonwords support the alternative vowel pronunciations more than others.

Method

Participants

Participants were 96 children in grades 2–5 from private and public schools. Demographic data on the sample are presented in Table 1. Sample raw and either scaled scores ($\bar{X}$ = 10, *SD* = 3) or standard scores ($\bar{X}$ = 100; *SD* = 15) for assessments of phonological awareness, rapid automatized naming, vocabulary, and general word reading skill, as well as set for variability and visual statistical learning (raw scores only) are provided in Table 2. In this study, we oversampled children who were struggling to learn to read words as represented by the depressed age-adjusted scaled and standardized scores representing phonemic awareness (scaled score=8.08), rapid naming (scaled score=8.10), and word reading (standard score=81.07). Although we oversampled for poor reading skills, the sample had normal age-adjusted scaled scores in vocabulary (scaled score = 10.08).

[1]In this study we use the terms “higher frequency” and “dominant” to refer to the unconditionalized vowel pronunciation (i.e., context-independent) and “lower frequency” and “nondominant” to refer to the conditionalized vowel pronunciation (i.e., context-dependent vowel).

Procedures

Testing occurred in the spring of grades 2–5, with trained research assistants administering all tests. All research assistants received extensive training and practice and were required to achieve 80% procedural fidelity before testing participants. Any discrepancies in administration were resolved, and support was provided throughout data collection in group and individual sessions. All testing sessions were audio recorded for scoring and reliability purposes. All tests were double scored and double entered by a fellow research assistant, and discrepancies resolved by the project coordinator. Twenty percent of test sessions for each tester were randomly selected to evaluate inter-rater reliability with reliabilities ranging from 97 to 100% ($\bar{X}$ = 98%).

Child Measures

Experimental nonword list—The dependent variable, an experimental measure of isolated nonword reading (N=76), comprised items sampled from the Treiman et al. (2006) list (n=28) along with additional items (n=48) developed to assess children's sensitivity to variant vowel pronunciations (for the complete list see Appendix A). This list included nonwords in which vowel GPCs varied in terms of whether the more and less frequent vowel GPC was wholly supported by the rime context (e.g., *cheam* and *chold*, respectively) and whether the more and less frequent GPC was favored by the rime context (e.g., *sint* and *drook,* respectively). In all cases, we use the same onset (e.g., *wook* vs. *woon*) across critical nonword comparisons to control for onset difficulty. As mentioned, two separate coding schemes were used: In the HF-GPC coding scheme, nonword pronunciations were scored as correct if the more frequent vowel GPC pronunciation was used (*chead* rhyming with *bead*), whereas in the LF-GPC coding scheme responses were considered correct if the less frequent vowel GPC pronunciation was used (*chead* rhyming with *head*). This was done because three possible responses were feasible at the item level (i.e., higher frequency GPC use, lower frequency GPC use based on the rime, and other) and allowed us to contrast child-level and nonword-level predictors across the two coding schemes. Although the two item-level prediction models based on HF-GPC and LF-GPC can be considered independent, there is dependency across coding schemes because only a single response by participants can be given for each nonword.

Phonemic awareness—The phonemic awareness task was the Elision task from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, Rashotte, & Pearson, 2013). Students were asked to delete phonological units from words. The authors report test-retest reliability of .93 (Wagner, Torgesen, Rashotte, & Pearson, 2013).

Rapid automatized naming—To test for rapid automatized naming, we used the letter naming task from the CTOPP (Wagner, Torgesen, Rashotte, & Pearson, 2013). For this task, students were asked to name a series of letters as fast as they could without making mistakes. The total score was the number of seconds students took to name all of the letters. Test-retest reliability is .72 for children of ages 8–17 years according to the test manual (Wagner et al., 2013).

Set for Variability (SfV)—Based on the work of Tunmer & Chapman (1998; 2012), SfV was assessed by examining participants' ability to determine the correct pronunciation from spoken words that were "mispronounced", as they might be if they were regularized or partially decoded (e.g., "breekfast" for *breakfast*). For this task, students were told they were going to play a word game with Alex and they were asked to figure out what Alex was trying to say. They were given two practice items ("/mŏther/" and "/brēkfăst/") with corrective feedback before they began the task. This task was administered with an audio recording. Kearns, Rogers, Koriakin, and Al Ghanem (2016) report a coefficient alpha of .82 for this task in children ages 7–11 years.

Vocabulary—The vocabulary subtest from the WASI (Weschler, 2011) was used to measure expressive vocabulary. The test requires students to identify pictures and define words. Interrater reliability for elementary age children range from .92–.94 (McCrimmon & Smith, 2013).

Visual statistical learning (VSL)—The VSL task used was based on previously published tasks (most notably, Siegelman, Bogaerts, & Frost, 2017; Arciuli & Simpson, 2012; 2011). The VSL task has two components: an exposure phase and a test phase. In exposure, twelve abstract shapes were presented one-by-one in a continuous stream at the center of the display for 550 ms each. Shapes were always presented in triplets and each triplet was presented 30 times. Participants were asked to press the spacebar when they saw a repetition of two shapes in a row. Participants were not told of the patterns. After exposure, participants were presented with 32 two-alternative forced-choice trials. After the triplets were presented, participants were asked to identify which of the triplets appeared in exposure. Participants then completed 16 two pattern completion tasks where two shapes from a triplet were presented and participants were asked to complete the pattern (a three-alternative forced-choice task). A combined score of 21 (43.75%) on the 48 item task is considered statistically above chance performance ($p < .05$).

Sight word reading efficiency—The word reading task in this study was the Sight Word Efficiency task from the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 2012). Students were asked to read a list of words in order of difficulty for 45 seconds. The maximum score is 108 and the authors report an alternate forms reliability of . 91.

Word measure

Rime type—To capture the frequency with which a lower frequency pronunciation of a vowel GPC is supported by the corpus of rime patterns (e.g., *oo* pronounced as /ʊ/ in *–ook*) in English, we calculated the proportion of words containing the target rime that corresponded with the lower frequency vowel GPC over the total number of words containing the rime using the Children's Printed Word Database (see Masterson, Stuart, Dixon, & Lovejoy, 2010). All counts were based on word types, not weighted by frequency in running text, consistent with the method employed by Kessler and Treiman (2001). The range for this measure was 0–1 ($\bar{X}$=.37, SD=.40). This measure is considered a proxy for the strength of the rime pattern to support the lower frequency pronunciation of the GPC in the nonword (i.e., context-dependent vowel pronunciation).

Data analysis

Missing data rates for the observed measures ranged from 2% on the sight word reading and phonological awareness measures to 12% on the VSL measure. Little's test of data missing completely at random (MCAR) resulted in a non-significant effect [$\chi^2(67) = 72.51$ p=.301] suggesting that the data were MCAR. As such, either multiple imputation or full information maximum likelihood could be used in model estimation. Because the student data were a component of the explanatory item response model (EIRM), and such models do not use a full information estimator, we opted to use multiple imputation. The multiple imputation was conducted with 1,000 imputations and values were aggregated across the imputations to construct a single-level student file to be used in the EIRM.

EIRM models allowed us to partition the item-level variance across children and nonwords. Random intercepts were included for child and nonword. Fixed effects were included for all child-level features, nonword-level features, and a child x nonword interaction (i.e., word reading x rime support). Separate EIRM models were generated for the unconditionalized HF-GPC and the conditionalized, rime influenced, LF-GPC coding schemes. The unconditional model (Model 0) was fit first by adding a person-specific random intercept (r_{010j}) and an item-specific random intercept (r_{020i}) because we expected random variation related to each of these variables. The binary outcome (p_{ji}; the probability of a correct response from person j on item i) was assumed to follow the Bernoulli distribution and random effects were assumed to be normally distributed. We used an unconditional model with only random effects for persons and items to determine the variability associated with persons and items. This model allowed us to determine how well the subsequent models explained this variance. Next, we added fixed effects for the aforementioned child and nonword characteristics (Model 1). Finally, we added exploratory interaction terms to explore our final research question (Model 2). No random slopes were included for these predictors.

We estimated the variability explained by calculating the reduction in child and nonword variance from the base model using the formula $(r_{010(Base\ model)} - r_{010(model\ n)})/r_{010(Base\ model)}$, where n represents the model to which the base model was compared (Bryk & Raudenbush, 1992). A detailed description of these analyses is beyond the scope of this report, but these models have been widely reported in the literature (e.g., Duff & Hulme, 2012; Gilbert, Compton, & Kearns, 2011; Goodwin, Gilbert, & Cho, 2013; Kearns et al., 2015; Kim, Petscher, Foorman, & Zhou, 2010; Steacy et al., 2016; Steacy et al., 2017).

Results

Zero order correlations between correct responses based on the two coding schemes (HF-GPC and LF-GPC), child variables, and the overall nonword base rate of success on the experimental nonword task (providing any acceptable answer) are provided in Table 3. Differences existed between the correlations of the overall nonword base rate with the rates of correct responses based on HF-GPC (.98) and LF-GPC (.81), although these differences could not be tested statistically. This was expected given the nature of the scoring schemes, with HF-GPC giving credit for more "traditional" use of GPC relationships. The opposite relationship, favoring LF-GPC, was present when considering correlations among HF-GPC (.43), LF-GPC (.62), and sight word reading. Additionally, there were significant correlations between all child-level predictors and nonword base rate performance, with the exception of the visual statistical learning task. The child level predictors were correlated with nonword base rate ranging from .28 (WASI Vocabulary) to .71 (Set for Variability).

As mentioned, separate EIRM models were developed for each of the coding schemes (i.e., HF-GPC and LF-GPC) employing the same child and nonword predictors. Overall, results indicate that there was significant item variance associated at the level of the child and nonword for both coding schemes (see Model 0 for HF-GPC and LF-GPC, Table 4). The unconditional model for HF-GPC had a logit intercept of .41 indicating that in the absence of child and nonword predictors students had a .61 probability of reading the nonword correctly using the most frequent vowel GPC, whereas the unconditional model for LF-GPC had a logit intercept of −1.80 indicating that in the absence of child and nonword predictors students had a .14 probability of assigning a nonword the alternative pronunciation. The student intraclass correlation (ICC) is .65 for HF-GPC compared to .22 for LF-GPC, whereas the nonword ICC is .35 for HF-GPC compared to .79 for LF-GPC, indicating differential breakdowns of variance align to the coding schemes.

Considering main effects first (see Models 1 for HF-GPC and LF-GPC, Table 4), in model HF-GPC there was a significant main effect for child phonological awareness (γ =.07, z=3.35), set for variability (γ=.05, z=4.17), and a significant main effect at the word level for rime type (γ =−1.68, z=5.57). For the LF-GPC model there was a significant main effect for child set for variability (γ=.02, z=2.27), sight word efficiency (γ=.03, z=6.01), and a significant main effect at the word level for rime proportion (γ =3.20, z=6.89). Rime proportion was negatively related to the probability of reading the nonwords correctly according to HF-GPC and positively related to the probability of reading the nonwords correctly according to LF-GPC, consistent with the rime proportion coding scheme.

Adding the exploratory interaction term between child word reading skill and rime type indicates that the relationship between rime type and nonword reading is moderated by reading skill in both models. The results of the interaction models are presented in Table 4 (Model 2) and are graphed in Figures 1 and 2. These results indicate that rime proportion has a greater impact on correct nonword pronunciation according to HF-GPC and LF-GPC for students with higher reading skill than for students with lower reading skill.

Discussion

The current study was designed to replicate and extend the Treiman et al. (2003; 2006) studies by examining child- and nonword-factors that help to explain trade-offs in developing readers use of unconditionalized vowel GPCs versus contextualized relationships between vowel and coda when reading nonwords with a variant vowel. The results suggest that general reading skill at the child-level and rime support at the nonword-level uniquely facilitate context-dependent vowel pronunciations in our sample of developing readers. We interpret the results within a statistical learning framework as supporting a developmental model of word reading in which children are more likely to use an alternative vowel pronunciation in a nonword as they become more proficient readers and as the occurrence of the alternative vowel pronunciation is increasingly supported by the corpus of words. As would be expected in such a model of word reading development, child and corpus attributes work to tune variant vowel pronunciations across individual children and nonwords, with important variance associated with both factors.

At the level of the child, our results indicate that the probability of students using the more frequent vowel GPC was higher in children with elevated phonological awareness, SfV, and word reading skill (see Table 4: HF-GPC, Model 2 & 3). We interpret these results as supporting a self-teaching model in which stronger phonological and word reading skills support children's use of the most frequent vowel GPC when encountering a new word. Alternatively, the probability of students using the less frequent GPC was associated with superior SfV and word reading skills (see Table 4: Code 2, Model 2 & 3). These results are similar to those reported by Treiman et al. and also support a self-teaching model of word reading development, albeit a more nuanced version, whereby increased reading skill improved the probability of using the conditionalized vowel pronunciation. We interpret these results as supporting a developmental model in which developing readers initially form simple one-to-one correspondences, based on high frequency GPC relationships, that are relatively insensitive to orthographic context (see Share, 1995) but through reading experience the O-P relationships become more sophisticated and context-dependent connections (Treiman et al., 2006). These results are expected, given that using the alternative, and less frequent GPC, pronunciation depends largely on students' exposure to these alternative pronunciations through reading and perhaps on a more flexible use of O-P relationships when pronouncing new words that comes with reading experience.

At the nonword level, a predictor of correctly pronouncing the nonword (with either the dominant or nondominant vowel GPC) was the corpus-based rime support for the particular vowel pronunciation. The direction of this relationship differed depending on whether the nonword vowel was scored for the more frequent GPC (i.e., HF-GPC: unconditionalized) or the less frequent vowel GPC (i.e., LF-GPC: conditionalized) pronunciation. In the case of the most frequent vowel GPC pronunciation (HF-GPC), the proportion of words with a rime that supported the conditionalized rime pronunciation negatively predicted the probability of reading the nonwords correctly. This finding suggests that the more words there are in the corpus containing a rime supporting the conditionalized vowel pronunciation, the less likely students were to read the nonwords using the most frequent unconditionalized vowel GPC. Alternatively, in the case of the conditionalized vowel pronunciation (LF-GPC), the proportion of words containing a rime supporting the conditionalized vowel pronunciation positively predicted the probability of the students reading the nonword correctly. This finding suggests that the more words there are in the corpus containing the alternate pronunciation of the rime, the more likely students were to produce the less frequent pronunciation of the vowel GPC. These results are consistent with what we would expect based on exposure to a broader corpus of words and support the role of statistical learning in word reading development. The significant interaction terms in Model 2 further support the role of exposure to a broader corpus in vowel pronunciation, with better readers demonstrating more sensitivity to the statistical regularities in the rime unit. These results are also consistent with Kessler's (2009) speculation that children are more likely to tune into orthographic context when there is no candidate that clearly dominates across contexts.

These corpus level effects are further supported by examples from the sample of nonwords used in this study. Using HF-GPC, the words that had the lowest number of correct responses in our sample of readers were *jook*, *grook*, and *drook.* These nonwords all share a rime unit with the high frequency word *book* and contain a rime unit that has a high proportion of the alternative rime pronunciation (i.e., only the word *spook* has the dominant /u/ pronunciation of *oo* before /k/). Again, this result is quite similar to that reported by Treiman et al. (2003). Alternatively, using LF-GPC, the nonwords that had the lowest number of correct responses in our sample were: *pont*, *slint*, and *dut.* These nonwords have a relatively low proportion, often a singular example, of the alternative vowel pronunciations: *front*, *pint*, and *put*, respectively. The case of *put* is an interesting one because it is a high frequency word that contains the alternative pronunciation. This vowel pronunciation may be a special case given the subtle difference between the pronunciation of *u* in *nut* and *put*, and the strength of the *u* pronunciation as /ʌ/ within the general corpus.

Our results diverge slightly from those reported by Treiman et al. (2006), in that the proportion of items in which the alternative pronunciation was used by children was lower in our sample (an overall probability of 14% across children and nonwords). We suggest that these differences can be attributed to two differences between the experiments: (1) our study oversampled for students with lower reading skills and (2) many of the nonwords in our sample had a proportion of alternative rime pronunciations of zero (e.g., *bimp* & *joom* - where there were no words in which the rime promoted the alternative pronunciation of the vowel), similar to Treiman et al.'s control nonwords. The restricted range of reading skills and higher proportion of nonwords that favor the use of the dominant GPC could have impacted the resulting probability estimates.

Finally, we would like to comment, and further speculate, regarding several unexpected relationships between SfV, VSL, and nonword vowel pronunciation across the HF-GPC and LF-GPC models. SfV was an important child-level predictor across both HF-GPC and LF-GPC models. In fact, it was the strongest predictor of nonword reading base-rate (the proportion of acceptable answers to the target nonwords) surpassing both phonemic awareness and word reading efficiency (see Table 3). It was unanticipated that SfV would be a stronger predictor of individual differences in nonword reading across the two coding schemes compared to other child-level predictors such as phonemic awareness skill, and this made us wonder exactly what skills the SfV task taps. We suggest that there are several skills involved in performance on the SfV task that are related to word reading. We agree with Kearns et al.'s (2016) hypothesis that the SfV task "measures a process that allows readers to take the output of phonological recoding assembled using phonological awareness skills and test it against entries in the phonological lexicon using lexical and sublexical semantic knowledge" (p. 457). However, we believe there may be more involved, hypothesizing that during the process of reading an irregular word, the structure of the item's phonological representation evolves to better reflect actual O-P relationships (for a detailed discussion see Elbro & de Jong, 2017). Thus, we speculate that there is an orthographic component (associated with reading experience) to the SfV task, moving beyond mere phonological assembly and lexical testing, which links this task to word and nonword reading skill above and beyond the level of straight phonological processing.

Results also indicate a stronger relationship between SfV and nonword reading performance in the case of HF-GPC compared to LF-GPC scoring schemes (although differences across models could not be tested statistically). This disparity in relationships across models is likely due to the strong phonological requirements of the SfV measure and the differential correlations between HF-GPC, LF-GPC, and nonword base-rate performance. Specifically, the reliance of the HF-GPC scoring scheme on the application of high frequency vowel GPCs resulted in higher correlations with nonword base-rate and all phonologically-based child-level predictors (i.e., SfV and phonemic awareness) indicating a scoring scheme that puts a premium on phonological decoding skills. In addition, the oversampling for poor readers who likely had more experience and training on HF-GPCs than LF-GPCs may have further strengthened the already strong relations between SfV and HF-GPC performance.

The second unexpected result was the lack of a relationship between VSL and nonword reading in both HF-GPC and LF-GPC models. We certainly expected VSL performance to be associated with the higher use of the conditionalized vowel in the LF-GPC model. We can think of a few possible explanations to explain this result which goes against a growing set of studies linking VSL type tasks and reading development (for a review see Arciuli & Simpson, 2012). The first explanation has to do with our sample. In oversampling for poor word readers in our study we may have truncated the range of statistical learning and nonword reading performance thus reducing the chance of detecting a relationship. The second could be our reliance on nonword reading as the outcome measure, as exemplified by the significant correlation between sight word reading and VSL displayed (see Table 3). Finally, we suggest it might be overly simplistic to propose a literal and causal connection between domain-general measures of statistical learning (e.g., VSL) and the domain-specific statistical learning routines that result in children considering the alternative vowel pronunciation in nonwords. A more probable model is one in which a set of general processes link domain-general statistical learning with word reading development (see Sawi & Rueckl, this issue), but the overlap does not include specific skills that relate word reading experience to individual differences in children’s sensitivity to the role that the rime coda plays in vowel pronunciation. Compounding the problem is the fact that domain-general measures of statistical learning tend to be difficult to measure reliably in children (see West, Vadillo, Shanks, & Hulme, 2017) and are not specifically designed to detect individual differences in statistical learning (e.g., Reber, Walkenfeld, & Hernstadt, 1991; Siegelman, Bogaerts, & Frost, 2017). Thus, it’s possible that the VSL task used in the study was not sensitive to the statistical learning skills of the students. As work continues examining the relations between statistical learning and reading development more consideration for the effect of sample, statistical learning task, and outcome measure will be important.

Overall, our results support a model in which a different set of child-level and word-level predictors are associated with the use of more and less frequent vowel GPC pronunciations and suggest that reading skill and rime support facilitate context-dependent vowel pronunciation in developing readers. The results of both the child-level and word-level predictors indicate corpus level effects that support the important role that statistical learning plays in students' reading of novel nonwords with variant vowel pronunciations. Our study was conducted in English but we speculate that similar results would be found in other complex, quasiregular orthographies such as French. Work done by Senechal (e.g., Senechal, Gringas, & L'Heureux, 2016) and Pacton (Pacton, Fayol, & Perruchet, 2005; Pacton, Perruchet, Fayol, & Cleeremans, 2001) suggests that French speaking students pick up on the statistical properties of the French orthography. More work on children's sensitivity to these regularities across orthographies will be important.

Acknowledgments

This research was supported in part by Grant P20HD091013 from NICHD. The content is solely the responsibility of the authors and does not necessarily represent the official view of NICHD.

References

Andrews S, Scarratt DR. 1998; Rule and analogy mechanisms in reading nonwords: Hough dou peapel rede gnew wirds? Journal of Experimental Psychology: Human Perception and Performance. 24(4): 1052–1086.

Arciuli J, Simpson IC. 2012; Statistical learning is related to reading ability in children and adults. Cognitive Science. 36:286–304. [PubMed: 21974775]

Bryk, AS, Raudenbush, SW. Hierarchical linear models for social and behavioral research: Applications and data analysis methods. Newbury Park, CA: Sage; 1992.

Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. 2001; DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychological Review. 108:204–256. [PubMed: 11212628]

Duff FJ, Hulme C. 2012; The role of children's phonological and semantic knowledge in learning to read words. Scientific Studies of Reading. 16:504–525.

Daniels DT, Share DL. Writing system variation and its consequences for reading and dyslexia. Scientific Studies of Reading.

Ehri LC. 2014; Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. Scientific Studies of Reading. 18(1):5–21.

Elbro, C, de Jong, PF. Orthographic learning is verbal learning. In: Cain, K, Compton, DL, Parrila, RK, editors. Theories of Reading Development. Amsterdam, The Netherlands: John Benjamins Publishing; 2017. 169–189.

Frost R. 2012; Towards a universal model of reading. Behavioral and Brain Sciences. 35:263–279. [PubMed: 22929057]

Gilbert JK, Compton DL, Kearns DK. 2011; Word and person effects on decoding accuracy: A new look at an old question. Journal of Educational Psychology. 103:489–507. [PubMed: 21743750]

Glushko RJ. 1979; The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance. 5(4):674–691.

Goodwin AP, Gilbert JK, Cho S. 2013; Morphological contributions to adolescent word reading: An item response approach. Reading Research Quarterly. 48(1):39–60.

Johnson DD, Venezky RL. 1976; Models for predicting how adults pronounce vowel digraph spellings in unfamiliar words. Visible Language. 10(3):257–68.

Kearns DM, Rogers HJ, Koriakin T, Al Ghanem R. 2016; Semantic and phonological ability to adjust recoding: A unique correlate of word reading skill? Scientific Studies of Reading. 20(6):455–470.

Kearns DM, Steacy LM, Compton DL, Gilbert JK, Goodwin AP, Cho E, … Collins AA. 2016; Modeling polymorphemic word recognition: Exploring differences among children with early-emerging and late-emerging word reading difficulty. Journal of learning disabilities. 49(4):368–394. [PubMed: 25331757]

Kessler B. 2009; Statistical learning of conditional orthographic correspondences. Writing Systems Research. 1:19–34.

Kessler B, Treiman R. 2001; Relationship between sounds and letters in english monosyllables. Journal of Memory and Language. 44(4):592–617.

Kim Y, Petscher Y, Foorman BR, Zhou C. 2010; The contributions of phonological awareness and letter-name knowledge to letter-sound acquisition—a cross-classified multilevel model approach. Journal of Educational Psychology. 102(2):313–326.

Masterson J, Stuart M, Dixon M, Lovejoy S. 2010; Children's printed word database: Continuities and changes over time in children's early reading vocabulary. British Journal of Psychology. 101(2): 221–242. [PubMed: 20021708]

McCrimmon AW, Smith AD. 2013; Review of Wechsler abbreviated scale of intelligence, second edition (WASI-II). Journal of Psychoeducational Assessment. 31(3):337–341.

Pacton S, Fayol M, Perruchet P. 2005; Children's implicit learning of graphotactic and morphological regularities. Child Development. 76:324–339. [PubMed: 15784085]

Pacton S, Perruchet P, Fayol M, Cleeremans A. 2001; Implicit learning out of the lab: the case of orthographic regularities. Journal of Experimental Psychology: General. 130:401. [PubMed: 11561917]

Perfetti C, Stafura J. 2014; Word knowledge in a theory of reading comprehension. Scientific Studies of Reading. 18(1):22–37.

Plaut DC, McClelland JL, Seidenberg MS, Patterson K. 1996; Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychological review. 103(1): 56–115. [PubMed: 8650300]

Reber AS, Walkenfeld FF, Hernstadt R. 1991; Implicit and explicit learning: Individual differences and IQ. Journal of Experimental Psychology: Learning, Memory, and Cognition. 17(5):888–896.

Ryder RJ, Pearson PD. 1980; Influence of type–token frequencies and final consonants on adults' internalization of vowel digraphs. Journal of Educational Psychology. 72(5):618–624.

Seidenberg, MS. Reading in different writing systems: One architecture, multiple solutions. In: McCardle, P, Miller, B, Lee, J, Tzeng, O, editors. Dyslexia across languages. Orthography and the Brain-Gene-Behavior Link. Baltimore, MD: Paul Brooke Publishing; 2011. 151–174.

Sénéchal M, Gingras M, L'Heureux L. 2016; Modeling spelling acquisition: The effect of orthographic regularities on silent-letter representations. Scientific Studies of Reading. 20(2):155–162.

Seymour PH, Aro M, Erskine JM. 2003; Foundation literacy acquisition in European orthographies. British Journal of psychology. 94(2):143–174. [PubMed: 12803812]

Siegelman N, Bogaerts L, Frost R. 2017; Measuring individual differences in statistical learning: Current pitfalls and possible solutions. Behavior research methods. 49(2):418–432. [PubMed: 26944577]

Steacy LM, Elleman AM, Lovett MW, Compton DL. 2016; Exploring differential effects across two decoding treatments on item-level transfer in children with significant word reading difficulties: A new approach for testing intervention elements. Scientific Studies of Reading. 20(4):283–295. [PubMed: 28596701]

Steacy LM, Kearns DM, Gilbert JK, Compton DL, Cho E, Lindstrom ER, Collins AA. 2017; Exploring individual differences in irregular word recognition among children with early-emerging and late-emerging word reading difficulty. Journal of Educational Psychology. 109(1):51–69.

Sawi O, Rueckl J. Reading and the neurocognitive bases of statistical learning. Scientific Studies of Reading.

Torgesen, JK, Wagner, R, Rashotte, C. Test of Word Reading Efficiency 2. Austin, TX: Pro-Ed; 2012.

Treiman R, Kessler B. 2006; Spelling as statistical learning: Using consonantal context to spell vowels. Journal of Educational Psychology. 98(3):642–652.

Treiman R, Kessler B, Bick S. 2003; Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. Cognition. 88(1):49–78. [PubMed: 12711153]

Treiman R, Kessler B, Zevin JD, Bick S, Davis M. 2006; Influence of consonantal context on the reading of vowels: Evidence from children. Journal of Experimental Child Psychology. 93(1):1–24. [PubMed: 16115645]

Tunmer, WE, Chapman, JW. Language prediction skill, phonological recoding ability and beginning reading. In: Hulme, C, Joshi, RM, editors. Reading and spelling: Development and disorder. Hillsdale, NJ: Erlbaum; 1998. 33–67.

Tunmer WE, Chapman JW. 2012; Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills? Scientific Studies of Reading. 16(2): 122–140.

Venezky, RL. The American way of spelling: The structure and origins of American English orthography. New York: Guilford Press; 1999.

Wagner, RK, Torgesen, JK, Rashotte, C, Pearson, NA. Comprehensive Test of Phonological Processing 2. Austin, TX: Pro-Ed; 2013.

Wechsler, D. Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II). San Antonio, TX: NCS Pearson; 2011.

West G, Vadillo MA, Shanks DR, Hulme C. 2017The procedural learning deficit hypothesis of language learning disorders: we see some problems. Developmental Science.

Ziegler JC, Goswami U. 2005; Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. Psychological Bulletin. 131(1):3–29. [PubMed: 15631549]

Appendix A: Experimental Nonword List

bimp
bink
brild *
brilt *
brold *
brond *
chead *
cheam *
chold *
chond *
crance *
crange *
dobe
dowth
drind *
drint *
drook *
droon *
dut
feam
foom
fow
fup
golt *
gont *
grook *
groon *

hean
hink
jead
jint
jook
joom
jowd
juff
keat
luff
nean
nind
plone
polt*
pont*
pove
powd
pown
sance*
sange*
slind*
slint*
sull
swead*
sweam*
swild*
swilt*
tobe
vead
veam
vind
voke
vone
voot
vown
vut
wook
woon
wup
zeat
zimp
zint
zoke

zoon
zoot
zove
zow
zowth
zull

*Indicates items from Treiman et al. (2006) study

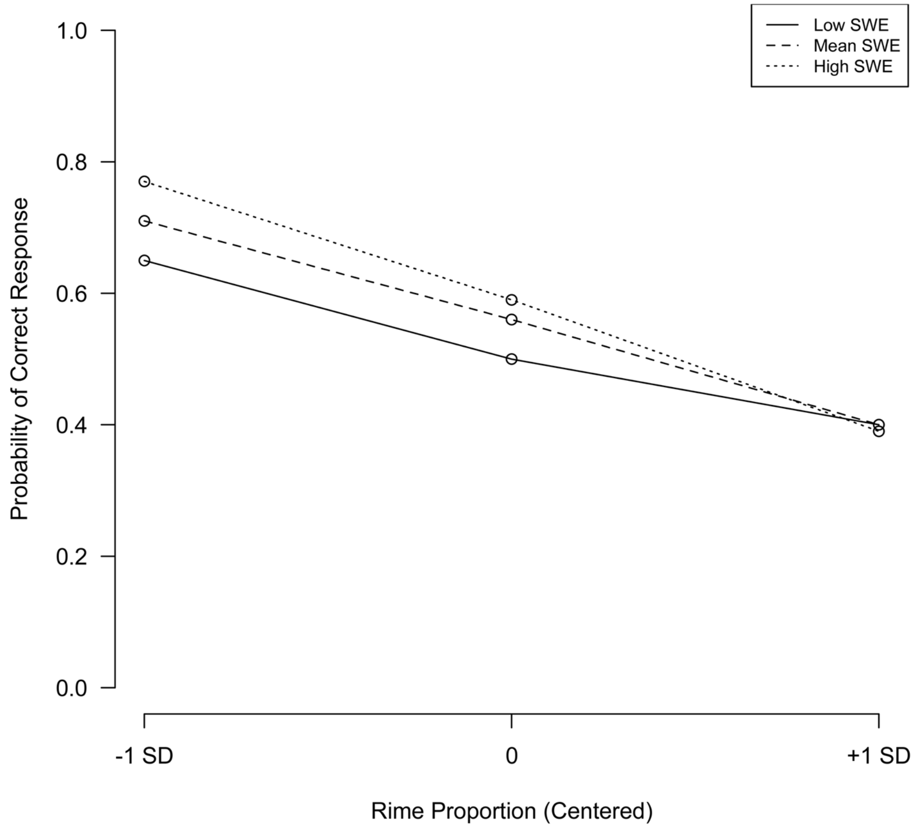

Figure 1.
Interaction between rime proportion and word reading using HF-GPC scoring scheme.

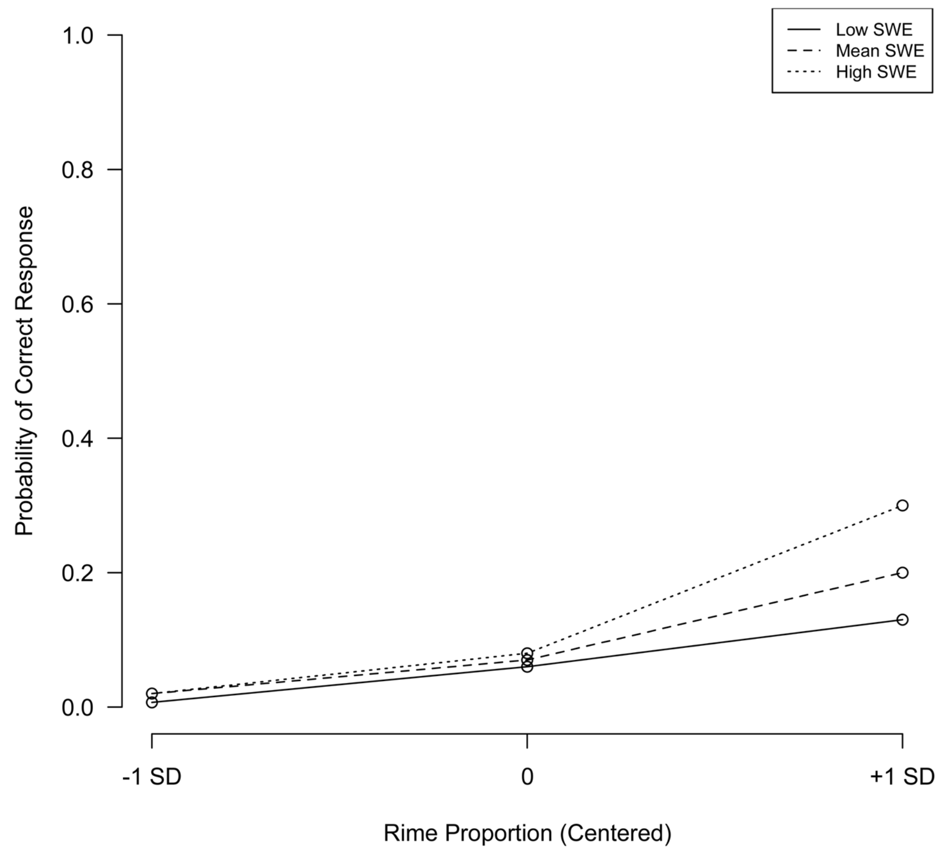

Figure 2.
Interaction between rime proportion and word reading using LF-GPC scoring scheme.

Table 1

Demographic Statistics

	Full Sample *N* = 96			
Variable	***n***	**%**	**Mean**	**(SD)**
Age (years)			9.91	(1.20)
Gender				
Female	43	44.79		
Male	53	55.21		
Race				
African American	16	17.20		
Hispanic	10	10.75		
Caucasian	65	69.90		
Multiracial	2	2.15		

Table 2

Child Level Descriptive Statistics

Variable	M	(SD)	Min	Max
Phonological Awareness				
Raw Score	22.06	(6.65)	0	33
Scaled Score	8.08	(2.90)	1	14
Rapid Automatized Naming				
Raw Score	21.48	(5.85)	12	41
Scaled Score	8.10	(2.40)	3	14
Set for Variability				
Raw Score	24.37	(11.10)	1	58
WASI Vocabulary				
Raw Score	25.80	(6.18)	1	37
Scaled Score	10.08	(3.25)	1	18
Visual Statistical Learning				
Raw Score (percent)	.49	(.10)	.31	.88
Sight Word Efficiency				
Raw Score	49.64	(17.22)	9	82
Standard Score	81.07	(15.29)	55	118

Table 3

Zero Order Correlations Between Child Variables

	1	2	3	4	5	6	7	8	9
1. Nonword Base Rate	–								
2. HF-GPC (Total)	**.98**	–							
3. LF-GPC (Total)	**.81**	**.71**	–						
4. Phonemic Awareness	**.63**	**.64**	**.41**	–					
5. Rapid Letter Naming	**−.42**	**−.35**	**−.38**	**−.32**	–				
6. Set for Variability	**.71**	**.67**	**.53**	**.67**	**−.33**	–			
7. WASI Vocabulary	**.28**	**.26**	.11	**.44**	−.18	**.37**	–		
8. Visual Statistical Learning	.12	.06	.05	−.02	−.09	.18	.12	–	
9. Sight Word Efficiency	**.58**	**.43**	**.62**	**.36**	**−.68**	**.50**	.13	**.20**	–

Note. Bolded numbers indicate significance at $p < .05$.

Table 4

Fixed Effects Models for Nonword Pronunciation

	Coding Scheme 1 (HF-GPC)						Coding Scheme 2 (LF-GPC)					
	Model 0		Model 1		Model 2		Model 0		Model 1		Model 2	
Fixed Effects	**EST**	**SE**	**EST**	**SE**	**EST**	**SE**	**EST**	**SE**	**EST**	**SE**	**EST**	**SE**
Intercept	.41*	.18	.23	.13	.25	.13	−1.80*	.22	−2.68*	.20	−2.60*	.20
Child Factors												
PA			.07*	.02	.06*	.02			.01	.01	.01	.01
RAN			−.02	.02	−.02	.02			.02	.01	.02	.01
SfV			.05*	.01	.05*	.01			.02*	.01	.02*	.01
Vocabulary			−.01	.02	−.01	.02			−.01	.01	−.01	.01
VSL			−.03	1.02	−.01	1.02			−.66	.64	−.67	.65
SWE			.01	.01	.01	.01			.03*	.01	.01	.01
Nonword Factors												
RP			−1.68*	.30	−1.67*	.30			3.20*	.46	2.99*	.47
Interaction												
SWE x RP					−.02*	.01					.06*	.01
Random Effects												
Child	1.76		.66		.65		.47		.12		.13	
Var. Exp.			.63		.64				.72		.72	
Nonwords	.94		.61		.62		1.72		.72		.76	
Var. Exp.			.35		.34				.57		.56	

Note. Est.= parameter estimate; SE=standard error; PA=phonemic awareness; RAN=rapid automatized naming; SfV=set for variability; Vocabulary=WASI vocabulary; VSL=visual statistical learning; SWE=sight word efficiency; RP=rime proportion; Var. Exp.=variance explained.

* $p<.05$