EFFECTS OF PREFERENCE AND STRATEGY ON LEARNING TO READ AN ARTIFICIAL SCRIPT

Jessica Wise¹, Yuliya Yoncheva², Bruce D. McCandliss²

¹Cognitive Studies and Psychology
²Psychology and Human Development
Vanderbilt University

ABSTRACT

Reading proficiency in the English alphabetic system requires mastery of common letter-to-sound mappings as well as rapid recognition of familiar words. Previous research has suggested that readers differ in their preferences for the two strategies of word reading, with some preferring to focus on the letter-to-sound correspondences within a word, a grapheme-phoneme (GP) strategy, while others preferring to match entire visual to spoken words, a whole-word (WW) strategy. The goal of this study was to assess what impact such preference differences might have on learning dynamics as readers learned a new writing system. First, typical adults’ reading preference for a GP vs. WW strategy in English was determined via a series of reading tests. Next, participants were trained to associate a novel script of artificial characters with spoken English words. Entire visual characters could be mapped onto whole spoken words (WW strategy) but if the hidden alphabet was taught, individual letters within the character could be mapped onto corresponding sounds (GP strategy). Participants in the two-day training learned one list of words using the GP strategy and another list using the WW strategy. It was hypothesized that preferences evident in reading English would elicit differences in learning dynamics for the novel script. Results revealed that participants’ preferences significantly predicted their learning patterns and outcomes. Specifically, WW preference learners suffered high transfer costs when applying letter knowledge to decoding novel words. Conversely, after learning new words, GP preference learners suffered from high interference costs on previously memorized words. These findings suggest that reading strategies may help uncover differences in cognitive challenges faced during reading acquisition, which in turn may inform effective teaching strategies in early reading instruction.

INTRODUCTION

The notion that children have different learning styles has had a long history in educational research (Honey & Mumford, 1992), especially in content areas such as early literacy acquisition. The construct of learning styles and strategies may provide insight into why some students excel at learning to read while others are merely mediocre or even lag behind. Is the difference something inherent within the student that allows them to excel at any type of learning or is the difference in response to the particular teaching method used? For instance, Baron and Strawson (1976) proposed two learning styles that correspond to the most commonly used methods for reading instruction in English. One is grapheme-phoneme (GP) in which readers map letters (graphemes) to sounds (phonemes) within spoken words. The second is whole word (WW) in which words are memorized as a complete unit. They proposed that readers could be classified based on how much of a preference a reader had for either of the two styles. Readers who relied more on the GP strategy were termed “Phoenician”, and those that relied more on the WW strategy were dubbed “Chinese” (referring to the Phoenician’s consistent alphabetic writing system, and the Chinese’s writing system of ideographs representing individual words). Phoenician readers were characterized by their superior knowledge of the pronunciation rules of English and were shown to be faster at reading lists of pseudowords (i.e., pronounceable non-words such as ‘smuncrit’) than lists of exception words (i.e., words that do not conform to conven-
tional pronunciation rules such as ‘colonel’). Chinese readers, conversely, were characterized by their ability to recognize the correct spelling of difficult to spell words and were shown to read lists of exception words faster than lists of pseudowords.

While numerous studies have employed the method of list reading to determine reading preference in order to examine the potential effects of preference on reading skills (e.g., Bowey & Ruth-erford, 2007; Treiman, 1984; Seghier, Lee, Schofield, Ellis, & Price, 2008; Freebody & Byrne, 1988; Bowey, 2008), the exact utility of such segregation is still a matter of debate. For instance, as readers become more experienced, preference groups might not show advantages in reading ability over other groups (Bowey, 2008). Further, segregating adult skilled readers into meaningful groups based on their potential preference has not always been successful (Brown, Lupker, & Colombo, 1994). Moreover, the actual advantage of having an extreme preference, (showing an imbalanced profile by leaning heavily toward a Phoeni-cian vs. Chinese preference) has also been challenged. In fact, Bowey and Ruthorford (2007) consider readers who show equal ‘preference’ to be the most balanced profiles in that they show no deficits on any reading measures compared to the preference readers.

Notably, an aspect of this issue that has remained elusive is examining the contribution of an individual’s learning style on new learning. Given the important implications these effects may have on future teaching programs, it is important to consider whether, and how, underlying preference might bias the learning outcome when learning a writing system. However, because preferences would emerge as readers first learn their native orthography, it is difficult to separate experience effects from preference effects in children. Compounding the challenge is the fact that when first learning to read in elementary school, students might receive explicit instruction in one type of reading strategy or another. This may cause young readers to rely on strategies they are specifically taught, instead of those they have a natural inclination towards. In order to control for all these factors inherently confounding preference in a typical learning situation, previous re-search has trained skilled adult readers to read an artificial orthography (or made-up writing sys-tem) as a model system for isolating specific fac-tors relevant to reading acquisition (Bishop, 1964; Yoncheva, Blau, Maurer, & McCandliss, 2010). For instance, an artificial orthography has been used successfully in previous research to demonstrate differential behavioral outcomes based on specific training strategies (Bitan, Manor, Morocz, & Karmi, 2005; McCandliss, Schneider & Smith, 1997). In fact, relevant to studying potential differences in learning style preference, Yoncheva et al. (2010) designed a study to examine the differences between the two most prevalent reading instruction methods: emphasis on GP strategy vs. emphasis on WW strategy. Their artificial orthography created a unique symbol for each letter, and therefore, words were composed of specific letter parts, but symbols for each word could be also viewed as a whole word as well. This feature of the artificial orthography allowed for examining the effects of both a GP strategy and a WW strategy trained on the same set of characters. Furthermore, since the writing system was new to all participants, they were all at the same starting point in learning, and thus obtained the same amount of experience over the course of training.

**CURRENT STUDY**

The aims of the current study are to examine the validity of learning style subgroups, as defined by differential performance on various list reading tasks, by examining how different subgroups perform in a novel learning task that involves learning to read a new writing system. Previous studies have used artificial orthography to demonstrate how differences in instruction lead to learning differences in a between-subject design (Yoncheva, et al., 2010). In this study, we aim to use a similar technique, but investigate learning differences in a within-subject design: each participant will learn using both strategies. We hypothesize that preference readers will show some sort of benefit in learning in their preferred strategy and a disad-vantage in their unpreferred strategy relative to the opposite preference group. Given that they are adults, we expect all participants to be able to learn under each strategy to a sufficient degree, regardless of potential preference. Thus, group differences may not manifest as simply better or worse performance during training, but rather, during later tasks that are designed to capture the
type of skills optimal for one learning strategy but not the other. Therefore, we will examine differences in transfer ability and interference effects between different preference groups. Transfer ability is an important skill for being able to read unfamiliar words by applying letter knowledge to unfamiliar words. In this study, transfer ability will be tested using novel ‘transfer’ words that have not been trained, but can be read using the letter knowledge gained in training. In order to assess transfer ability, we will contrast performance on assessments of trained GP words to performance on assessments of Transfer words. Interference effects, on the other hand, can manifest during WW learning. These are exhibited when performance on previously memorized words suffers after additional word learning. It is important not to suffer from these effects when trying to learn an increasing number of exception words in WW learning. In order to examine interference effects, we will compare assessments of a list of words both prior to and after additional word learning.

Because Chinese readers rely more on memorization of trained characters, it is hypothesized that Chinese readers will demonstrate poor transfer ability while Phoenician readers will perform the same on assessments of GP trained words and transfer words. Second, because Phoenician readers rely more on decoding, they may suffer from interference effects when additional WW words must be memorized, while Chinese readers would show little to no effects.

**METHODS**

**Participants**

All participants were right-handed (as confirmed by Edinburgh Handedness Inventory), native English speakers with no reading disabilities, as determined by a battery of standardized assessments: Test of Word Reading Efficiency (TOWRE), Nelson Denny Reading Fluency, and Comprehensive Test of Phonological Processing (CTOPP). Participants were excluded if reading scores on the combination of standardized measures were below 15th percentile, performance was below 70% on the behavioral tests on trained characters on either Day 1 or Day 2 (2 participants), or did not complete Day 2 (5 participants). The final analysis included 29 participants. The North American Reading Test (NART) and the Phonemic Decoding (PD) portion of the TOWRE test were used to determine a participant’s preference. The NART is composed of a series of increasingly more infrequent exception words, therefore measuring use of the WW strategy, while the PD of the TOWRE is composed of increasingly longer pseudowords, measuring use of the GP strategy. However, the NART is designed for ages 18 and above, while the TOWRE is designed for ages 6-24, but uses the same entry point for all ages. Given this imbalance in age range of the tests, a preference score was obtained for each participant by subtracting errors on the first 30 items on the NART from errors on the last 30 items on the PD portion of the TOWRE test. Errors were compared in this manner to match for item difficulty. This method for determining preference ensured that a participant’s preference was relative to his own performance, not that of the overall group. Therefore, high (or low) scores alone on either of the tests did not determine preference in and of itself, and preference reflected a large difference in individual’s TOWRE versus NART scores, rather than the absolute magnitude of either test separately. Categories were created using a tertile split to ensure equal group size (optimizing statistical power) as well as consistent grouping of scores (i.e. all participants with a score of -1 were in the same group). The mean preference score was 0, and participants 0.5 standard deviations away from the mean were classified as Phoenician or Chinese readers, while the rest were considered Balanced readers. Participants with a positive score of 2 and above (more errors on the TOWRE than the NART) were classified as Chinese readers (10) and those with a negative score of -2 or below (more errors on the NART than the TOWRE) as Phoenician readers (10). There were no significant differences between the three groups on reading tests aside from those used to classify. Demographics and test scores for each group can be seen in Table 1.
Table 1

Demographic Profiles and Mean Standardized Test Performance Scores of Participants by Preference

<table>
<thead>
<tr>
<th></th>
<th>Phoenicians Readers (n = 10)</th>
<th>Balanced Readers (n = 9)</th>
<th>Chinese Readers (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference Rank</td>
<td>Mean  sd</td>
<td>Mean  sd</td>
<td>Mean  sd</td>
</tr>
<tr>
<td>Number of Males</td>
<td>6  3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>22.90  5.59</td>
<td>22.33  4.27</td>
<td>24.30  5.31</td>
</tr>
<tr>
<td>Years of College Edu-</td>
<td>2.85  1.51</td>
<td>2.33  1.44</td>
<td>2.83  1.17</td>
</tr>
<tr>
<td>cation</td>
<td>RS TOWRE PD</td>
<td>RS NART</td>
<td>RS TOWRE SWE PR ND Reading Fluency</td>
</tr>
<tr>
<td></td>
<td>57.70*  4.83</td>
<td>40.22  6.06</td>
<td>101.67  2.60</td>
</tr>
<tr>
<td></td>
<td>31.30*  8.87</td>
<td>35.56*  8.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>99.10  6.66</td>
<td>101.67  2.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>78.50  35.35</td>
<td>94.22  12.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.60  2.12</td>
<td>15.11  1.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>56.80  8.12</td>
<td>57.67  4.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>57.30  7.96</td>
<td>60.11  2.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>57.30  7.96</td>
<td>60.11  2.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>57.10  4.70</td>
<td>60.11  2.85</td>
<td></td>
</tr>
</tbody>
</table>

Note. RS = raw score; PR = percentile ranking; SS = scaled score. * p < .05

Stimuli

Participants were trained on an artificial script similar to that used by Yoncheva, Blau, Maurer, and McCandliss (2010). The script was composed of letter-like characters combined to make words, which were composed of 6 consonants (b, d, k, n, s, t) and 4 vowels (a, e, i, u). Participants were trained on 24 words that could be read using the letter characters (a GP strategy) and 24 words that could only be learned by memorizing the whole symbol (a WW strategy). To aid in the reading process and make the WW script dissimilar from the GP script, the letters in the GP characters were stacked horizontally, and the potential 'letters' in the WW characters were stacked vertically. In order to ensure that a WW strategy was used on these characters, the 'letters' within the written symbol were never consistently mapped to sounds within the words. Therefore, even if a participant attempted to use his letter knowledge and learn using a GP strategy, it would not be helpful (e.g., for one word, the visual symbol for 'b' might be paired with the auditory 'm' and on another with the auditory 'k'). Participants were trained on lists of 12 words at a time. Within each list, each consonant was in both the first and last position of a word twice, and each vowel appeared three times.

Procedure

Participants took part in a two day study, during which they were trained on words using the artificial script described above. On each day, participants were trained on two lists and were instructed to use GP strategy on one list and WW strategy on the other. Participants completed three alternating training sessions on each list. Stimuli lists were counterbalanced across participants to control for strategy, day, and training order. Therefore, across different participants each list was presented as both a WW and a GP strategy list as well as in both first and second position on Day 1 and Day 2 for a total of 8 orders. A training session consisted of seven repetitions of each word followed by a test on the trained words to track learning progress. In the GP training session, the test of trained words was followed by a test of 6 'transfer' words that participants were not trained on, but could read using the letter knowledge learned from training. This
ensured the acquisition of GP strategy on GP words (despite not being able to specifically prevent WW strategy) since transfer words could not be recognized using a WW strategy. On each training trial, a word was presented for 2000 ms followed by the corresponding auditory word which lasted approximately 600 ms. Participants were asked to sound out each letter during GP training and say the entire word during WW training during the trial. Each test trial presented a trained word for 225 ms followed by an auditory word. Participants were instructed to decide whether the symbol matched the auditory word or not. Each visual word was presented twice: once matched with its audio, and once mismatched. The foil for a target was a matched word from the trained list that shared at least one letter with the target word. In order to prevent learning from the test, foil pairs were always presented together resulting in 24 test trials.

In addition to training, on Day 1, participants completed a standardized testing session. This consisted of the Reading Fluency portion of the Nelson Denny, the NART, the TOWRE, the performance portion of the WASI, and the Non-Word Repetition test in the CTOPP.

Before training on Day 2, Day 1 learning was assessed with a retention test. The second training session then began with two lists with 24 new words following the same procedure as Day 1. Day 2 ended with a final assessment of all trained words and a list of new transfer words. Participants completed five groups of short tests (24 trials) on each list for a total of 25 minitests, five repetitions of each list. Each group presented the lists in a different order to minimize effects from practice and fatigue.

RESULTS

Standardized Tests

No significant differences between the three groups on any one test were found, except for those tests used in classification: the NART and TOWRE PD. In order to better compare performance across tests, scaled scores were converted to z scores. Z-scores for all verbal tests not used in classification (Nelson Denny, TOWRE SWE, and CTOPP) were then averaged together to get one verbal performance test score for each participant. There were no significant correlations between test scores (both single test scores and averaged z-scores) and preference ranking. However, when comparing the absolute value of the preference score, there was a significant negative correlation between extremity of preference and average verbal test z-score r(27) = -.603, p < .001. This shows that regardless of what the preference is, the stronger the preference, the poorer the results on standardized reading tests.

Learning Task

Accuracy and reaction times were measured after each training session such that each participant had three data points (s1, s2, and s3) for each word list. RTs for incorrect trials and outside two SD away from the participant’s mean were excluded. Accuracy and RTs were compared at each time point, as well as collapsed across sessions. Overall, there was a main effect of Day, with all participants regardless of preference performing faster on Day 2 than Day 1, F(1,28)  = 16.34, p<.001, reflecting a general practice effect. This improvement was equal across all tasks, as there was not an interaction between Day and Strategy F(2,17) = 1.62, p = 0.22. No significant accuracy differences emerged between groups based on strategy at any point during both days indicating that training was equally successful for all preference groups for both strategies.

Between group differences on the learning assessments are summarized in Figures 2 and 3. We assessed learning for Day 1 at two different time points: prior to Day 2 training during the retention test and after Day 2 training session during the final testing. Day 2 was assessed only during the final tests. Analysis revealed no accuracy differences between groups on any GP or WW tests, all Z’s (19) < 2.0, p’s > .05. RT differences revealed that Balanced readers were faster than both Chinese and Phoenician readers on all tests of WW, all Z’s (19) > 2.0, p’s < .05. Balanced readers were also faster than Chinese (but not Phoenician) readers on the post-training test of Day 1 words, Z(19) = 2.36, p = .018. Both Phoenician and Balanced readers performed significantly faster than Chinese readers on the final transfer test, Z(19) > 2.1, p < .05.
Figure 1. Reaction times in ms for each test taken over each of the three sessions over the two days. Decreasing RTs over session indicate successful learning in both strategies (red vs. blue) for each of the three preference groups (left, middle, right panel).

Figure 2. Accuracy for each assessment of learning for each of the three preference groups. The lack of significant differences indicates that all groups were able to learn under each strategy.
Transfer Ability and Interference Effects

In order to test for transfer ability, we compared tests of GP trained words against transfer words. In order to demonstrate the most robust effects, RTs and accuracy from the final test of GP Day 1 words were compared with the novel Transfer test also taken during final testing. We calculated a transfer cost for each participant by subtracting Transfer performance from GP performance such that a higher number indicated a higher cost to transfer knowledge to a new word than to read a trained one. There were no significant differences found between groups on transfer cost of accuracy, but there were in RTs. All participants showed a positive score indicating that regardless of preference there was a cost to transfer to new words compared to recognizing trained words, which was expected. Phoenician readers, however, had a significantly lower transfer cost than both the Balanced and Chinese readers, \( Z(19) > 2.0, p < .05 \). Balanced and Chinese readers, though, did not differ, \( Z(19) = .33, p = .74 \).

To examine interference effects that result from training on more words in the WW strategy, we compared accuracy and RT data from the WW retention test and the final test of WW Day 1. We calculated an interference cost by subtracting final test performance from Retention Test performance such that a higher number indicated a higher cost to performance. In this case, analysis revealed no differences in RT, but differences in accuracy. Phoenician readers had a negative score; they performed worse on WW trained words learned on the first day after learning more words. Chinese readers, conversely, had a positive score, improving in their performance. Balanced readers performed similarly on both tests. Though there was a significant difference between the cost scores between Chinese and Phoenician readers \( Z(20) = 2.47, p = .011 \), Balanced readers did not perform significantly differently from either Chinese or Phoenician readers.

Figure 3. Reaction Time data for each assessment of learning for each of the three preference groups. Chinese readers were the slowest group on the transfer test and Balanced readers were the fastest responders to WW words on all assessments.
DISCUSSION

The present results indicate that differences in reading preferences, as defined by list reading tasks, may lead to significant differences in the cognitive profile of new learning. The current study contrasted different aspects of reading tests of English to independently classify participants as having a preference for one learning style over another and demonstrated that this preference had a robust impact on the type of new knowledge that was acquired when learning to read a novel writing system. Results showed that having a preference in either direction had both a positive and a negative effect on learning in some way, while having a balanced profile led to steady, progressive learning throughout, relative to the others groups. Chinese readers had a higher cost when transferring letter knowledge to new words compared to Phoenician readers, and Phoenician readers had a higher interference cost on WW trained words compared to the Chinese readers. In order to decipher whether the differences in learning profile reflect an advantage or disadvantage for the preference reader, results can be compared to the Balanced readers that showed equal preference. Phoenician readers can be considered to have an advantage in transfer ability over the other groups since they had a lower transfer cost than both Chinese and Neutral readers. However, the interference effect results are less conclusive. Because neither preference group performed differently from the Balanced readers the current data do not answer whether Phoenician readers demonstrated a deficit, Chinese readers demonstrated an advantage, or possibly both. Current research in our lab is investigating whether the advantage is worth the disadvantage for the preference learners, or if being a Balanced reader is overall the more desirable position, since the differences between learning are not so drastic between the two styles. Although further research is still needed, the score profiles from this study lend support to the notion that it is better for overall reading performance to not have a preference.

Further research is also necessary to determine how having an imbalanced reading profile compares to that of a balanced one. This study used tests that were designed to be optimal for each of the major reading strategies, but not for an approach that is balanced between the two. Therefore, the current study cannot speak to exactly what strategy Balanced readers are applying when learning using different strategies; whether they are flexible between the two styles and choose the most optimal strategy for the particular word or using a combination of both strategies.
for each word. If Balanced readers were performing optimally during each strategy, it would be expected that their performance would be similar to the preference readers who showed an advantage. However, these results did not indicate that Balanced readers had any advantages, merely no disadvantages when compared to both preference groups. The implications of this are currently under investigation in our lab.

**Broader Implications**

This current study, as a piece of a larger research program, has important implications for improving educational practices, specifically in determining the best method to teach reading. Many might assume that the best way to teach a burgeoning reader is to first determine what strategy he prefers to use to read and to then emphasize that strategy while teaching. However, this study indicates that preference readers do not necessarily learn better in their preferred strategy as opposed to their unpreferred strategy. Rather, they excel in their preferred style and suffer in their unpreferred strategy. Since both strategies are necessary in the reading process, instead of trying to accommodate preference readers by teaching using their preferred strategy, it may be better to try to lead imbalanced reading profile readers to have a more balanced profile by instructing them in the opposite strategy. Further research is needed to see just how affected readers’ performance would be if allowed to use their preferred strategy to learn new words without any restraints on the reading strategy used such as those that were in place in the current study. If participants were allowed to use any strategy as desired, the deficits found for the preference readers in this study could be magnified. For example, if Chinese readers never used a GP strategy, perhaps they would completely fail to read transfer words instead of merely being slow to respond to them. This could also shed light on the relationship between balanced and imbalanced readers. The implications from this line of research could become important when readers progress in their education and are expected to learn increasing amounts of vocabulary, but receive less formal instruction in the classroom as to how to learn them, and increasingly rely on personal preference and available strategies instead.

**ACKNOWLEDGEMENTS**

This project was funded in part by the Vanderbilt Undergraduate Research Project grant, and was advised by Yuliya Yoncheva, Ph.D., and faculty advisor Bruce D. McCandliss, Ph.D.

**REFERENCES**


