JSLHR

Research Article

The Effect of Stimulus Variability on Learning and Generalization of Reading in a Novel Script

Jasmeen Adwan-Mansour[a] and Tali Bitan[b]

Purpose: The benefit of stimulus variability for generalization of acquired skills and knowledge has been shown in motor, perceptual, and language learning but has rarely been studied in reading. We studied the effect of variable training in a novel language on reading trained and untrained words.
Method: Sixty typical adults received 2 sessions of training in reading an artificial script. Participants were assigned to 1 of 3 groups: a variable training group practicing a large set of 24 words, and 2 nonvariable training groups practicing a smaller set of 12 words, with twice the number of repetitions per word.
Results: Variable training resulted in higher accuracy for both trained and untrained items composed of the same graphemes, compared to the nonvariable training. Moreover, performance on untrained items was correlated with phonemic awareness only for the nonvariable training groups.
Conclusions: High stimulus variability increases the reliance on small unit decoding in adults reading in a novel script, which is beneficial for both familiar and novel words. These results show that the statistical properties of the input during reading acquisition influence the type of acquired knowledge and have theoretical and practical implications for planning efficient reading instruction methods.
Supplemental Material: https://doi.org/10.23641/asha.5302195

The aim of the current study is to examine the effect of stimulus variability on learning to read in a novel script. Motor and perceptual learning studies have shown the benefit of variable training to learning and generalization (Douvis, 2005; Huet et al., 2011; Schmidt, 1975; Wulf & Schmidt, 1997; Yao, DeSola, & Bi, 2009). Although the effect of stimulus variability has also been shown in linguistic tasks such as grammar learning and speech perception (Clopper & Pisoni, 2004; Lively, Logan, & Pisoni, 1993; Martin, Mullennix, Pisoni, & Summers, 1989; Perry, Samuelson, Malloy, & Schiffer, 2010; von Koss Torkildsen, Dailey, Aguilar, Gómez, & Plante, 2013), very few studies have tested it in the context of reading acquisition (Apfelbaum, Hazeltine, & McMurray, 2013). The current study examined the hypothesis that when learning to read in a novel script, high variability of the input would facilitate learning of letters and letter combinations, thus improving generalization to new words.

[a]Learning Disabilities Department, University of Haifa, Israel
[b]Psychology Department, The Institute of Information Processing and Decision Making, University of Haifa, Israel

Correspondence to Tali Bitan: tbitan@research.haifa.ac.il

Editor: Sean Redmond
Associate Editor: Nicole Terry

Received July 20, 2016
Revision received February 7, 2017
Accepted March 23, 2017
https://doi.org/10.1044/2017_JSLHR-L-16-0293

Variability

The concept of variable training has been developed in the context of motor learning studies (Schmidt, 1975). It highlights the idea that high variability in trained tasks enhances cognitive flexibility and results in improved motor performance in novel contexts. Various studies in this field had demonstrated positive effects for variable training on learning and generalization (Douvis, 2005; Yao et al., 2009).

The benefit of variability on learning and generalization had also been shown in auditory and speech perception learning tasks such as auditory frequency discrimination (Amitay, Hawkey, & Moore, 2005) or categorization of novel sentences to different regional dialects following multiple talkers condition (Clopper & Pisoni, 2004). In the same vein, Japanese listeners trained to identify English /r/ and /l/ sounds under multiple talkers condition and with a high variability stimulus set showed better generalization to new tokens and new talkers (Lively et al., 1993). Furthermore, learning second language vocabulary with acoustically degraded stimuli was significantly enhanced by multiple talkers compared to a single talker (Sommers & Barcroft, 2011).

In the language domain, an advantage of variable training was found for learning nominal categories in children (Perry et al., 2010) and in various grammar tasks in adults. For example, Reeder, Newport, and Aslin (2013)

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

showed that adults learning a linguistic category in an artificial language benefited from being exposed to a large number of sentence contexts, resulting in improved identification of familiar and novel grammatical strings (Reeder et al., 2013). In the same vein, learning of noun gender categories in a novel language (Russian) was facilitated by input variability (a large number of root words) (Eidsvåg, Austad, Plante, & Asbjørnsen, 2015). Variable training also facilitates learning of novel stems and affixes in a new vocabulary (Tamminen, Davis, & Rastle, 2015), and artificial grammar learning (Grunow, Spaulding, Gómez, & Plante, 2006; von Koss Torkildsen et al., 2013). Recently, the benefit of variability for learning and generalization was also shown in a therapeutic context in children with specific language impairment learning the morphological structure of verbs (Plante et al., 2014).

Very few studies have examined the effect of variability on learning how to read. In adults, practicing mirror-reading of familiar words, a small set of words repeated more times resulted in better performance on trained words compared to a large set of words practiced only once (Ofen-Noy, Dudai, & Karni, 2003), but the transfer to untrained words was not reported. However, different results were found in a study that examined children learning to read in English. Apfelbaum et al. (2013) examined the learning process of six grapheme-phoneme correspondences (A, I, O, AI, EA, and OA) in first-grade English-speaking children. They manipulated variability by presenting the vowels in words with more versus fewer consonants (19 vs. 10) resulting in different numbers of consonant frames (57 vs. 21). The group trained under the variable condition identified the vowels better in both trained and untrained words. The authors proposed that the variable consonant frames (the irrelevant aspect of the stimuli) helped children focus and identify the critical vowel targets (the relevant aspect). One limitation of this study is using the English alphabetic code to which participants have prior exposure. In addition, because this study manipulated the variability of irrelevant aspects of the stimuli (consonants when learning vowels), it is still not clear whether high variability of task-relevant aspects of the stimuli (i.e., trained words) can also improve learning to read.

Reading

Reading involves mapping between orthography, phonology, and semantics, which can occur at multiple sublexical and lexical size units (Coltheart, Curtis, Atkins, & Haller, 1993; Frost, 1992; Frost, 1994; Ziegler & Goswami, 2005). Although mapping of large orthographic units, such as whole words, to phonology and semantics requires prior acquaintance with the word, decoding of smaller grain-size units, such as letters and letter clusters to their phonological representations enables reading of new unfamiliar words. The size of the units being mapped in the reading process depends both on the consistency of the script system and on the individual reader's stage in the process of reading acquisition (Ziegler & Goswami, 2005). Thus, when reading orthographies with consistent mapping of graphemes to phonemes, readers rely more on small grain sizes in early stages of reading acquisition, compared to orthographies with inconsistent grapheme-phoneme relations where beginning readers rely more on large grain-size units (Ziegler & Goswami, 2005). However, even in a relatively inconsistent orthography, such as English, children's knowledge of small orthographic units positively affects orthographic identification already at early stages of reading acquisition (Cunningham, 2006). Importantly, in addition to the characteristics of the orthography and of the individual reader, the size of units being mapped during reading may also depend on the learner's specific experience, which is affected by the reading instruction method, and the units it emphasizes (Bitan & Karni, 2003; Brennan & Booth, 2015; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998).

After decades of debates about the effectiveness of reading instruction methods that emphasize small versus large unit sizes in children (Foorman, 1995), firm evidence has accumulated for the critical role of small unit instruction in reading acquisition in alphabetic orthographies (Hulme & Snowling, 2013). These instruction methods focus on phonemic awareness and letter knowledge (Shapiro & Solity, 2008); systematic-phonics instruction (de Graaff, Bosman, Hasselman, & Verhoeven, 2009); explicit instruction of the alphabetic principle (Foorman et al., 1998); and phonological recoding by reading aloud (Kyte & Johnson, 2006). According to Share's self-teaching hypothesis (Share, 1995) word recognition is rooted in prior successful decoding of its letter-sound correspondences. A reader with basic abilities of letter-sound mappings and phonemic awareness can use the phonological decoding of a grapheme sequence as a self-teaching mechanism and generalize the knowledge to new words or wider contexts (Share, 1995). This approach emphasizes the importance of generalization of the acquired knowledge, which is the goal and the challenge of learning in any domain (Perkins & Salomon, 1988; Schmidt & Bjork, 1992). There is empirical evidence that children implement both lexical and sublexical reading strategies and that alternating between large and small grain-size units benefits reading speed (Díaz et al., 2009). It has been shown that combining bottom-up training (emphasizing small units like graphemes) and top-down training (emphasizing large units like sentences and utterance) had positive effects on literacy skills (Helland, Tjus, Hovden, Ofte, & Heimann, 2011).

The current study focused on adults learning to read a novel orthographic system. Literate adults, familiar with the alphabetic principle, do not necessarily benefit from small unit instruction to the same extent as children learning to read in their native language. However, studies on adults learning to read in an artificial orthography (Bitan & Booth, 2012; Bitan & Karni, 2003, 2004; Bitan, Manor, Morocz, & Karni, 2005; Brennan & Booth, 2015) show the effect of reading instruction on the size of the acquired reading units. Specifically, directing participants' attention to the individual letters by prior exposure to letters before training on whole words results in greater reliance on letter decoding (Brennan & Booth, 2015) and better generalization to untrained words composed of the same letters (Bitan & Karni, 2003, 2004).

The Current Study

The current study tested the hypothesis that the statistical properties of the trained stimuli in early stages of exposure affect the size of the reading units. Specifically, we predicted that a larger set of trained words would increase the variability of the orthographic and phonological contexts in which each letter appears. This is expected to tune participants' sensitivity toward the individual letters, which are repeated across words, and away from whole-word patterns, thus increasing the probability of learning and mapping individual letters to their corresponding sounds. This, in turn, is expected to improve reading of untrained words in the novel script.

This prediction was tested using an artificial orthography, a paradigm which has been previously used in investigating the acquisition of reading skills (Bitan & Karni, 2004; Mei et al., 2013; Taylor, Plunkett, & Nation, 2011; Yoncheva, Blau, Maurer, & McCandliss, 2010). Artificial orthographies and languages are well-suited paradigms for examining learning and generalization because one can tightly control the statistical properties of the input. They are composed of a small set of items that can generally be learned to reasonably high proficiency over the course of hours. Importantly, results of studies using artificial languages and orthographies indicate that participants' performance on the artificial system are positively correlated with measures of acquisition of natural languages and orthographies (Ettlinger, Morgan-Short, Faretta-Stutenberg, & Wong, 2015; Plante et al., 2014; Taylor et al., 2011). Moreover, our previous studies, using the same artificial orthography used in the current study had shown that participants' learning is correlated with the standardized reading scores in their native language (Bitan & Booth, 2012) and activates the reading network in the brain (Bitan et al., 2005).

In the current study, we examined the effect of variability on learning to read an artificial script and on generalization to new words. Variability was manipulated between groups by changing the number of words in the trained set (12 vs. 24 words), while keeping the number of different letters they contain and the total number of practice trials constant across groups. Given previous evidence for the advantage of variable practice for generalization to untrained items, we predicted that high variability (a large set of words) would hinder learning of whole word units and thus facilitate learning of the individual letters, compared to a small set of trained words. Thus, we expected that variable training would decrease performance on trained words but improve generalization to new words composed of the same letters. Such a pattern of results would not only expand the theory of variable practice (Schmidt, 1975) to reading, but would also suggest a mechanism by which it operates. The practical implications of these results could contribute to the design of effective reading instruction methods.

Finally, because we expected participants trained in the variable condition to rely more on decoding of small units, whereas participants trained in the nonvariable condition would rely more on memorization of whole words, we expected differential correlations with measures of phonemic awareness and working memory. Phonemic awareness skills are critical for decoding of small units (Meschyan & Hernandez, 2002). However, it is not clear whether phonemic awareness would be more important when training conditions facilitate decoding (i.e., in the variable training group) or when training conditions do not facilitate decoding (i.e., in the nonvariable training group), in which case decoding depends mainly on the individual's phonemic awareness.

Method

Participants

Sixty adults were recruited from undergraduate students at the University of Haifa who are typically from middle to high socioeconomic status. All were native speakers and schooled in Hebrew, with only exposure to foreign languages through formal instruction. They had no neurological or psychiatric disorders and reported normal hearing and vision. Informed consent was obtained before the experiment. To ensure all participants were typical readers without reading impairment, we excluded four participants due to low performance on a word reading test (± 1.5 *SD* from the mean of normal reading scores $N = 93$). One participant was excluded due to a technical error. Participants were randomly assigned to one of three groups, with 18 participants (five men, 13 women, mean age 25.16 ± 2.47) in the variable training group, 19 participants (five men, 14 women, mean age 25.57 ± 1.84) in the Nonvariable 1 group, and 18 participants (five men, 13 women, mean age 25.11 ± 2.10) in the Nonvariable 2 group.

Materials

Screening Tests

Working memory, word reading ability, object naming, and phonemic awareness were tested to ensure participants performed at the normal range and that the three groups were comparable in their abilities. Because working memory and phonemic awareness could potentially contribute to performance in the experimental tasks, they were also used as covariates in the statistical analysis.

Working memory was tested using the digit span forward and backward task age-appropriate standard scores (Wechsler, 1999). Reading ability was tested using the words per minute (Shatil, 1995b) and nonwords per minute (Shatil, 1995a) tests. Phonemic awareness was tested using the phonemic deletion test (Ben Dror & Shani, 1996), and naming fluency was tested using the object naming task (Denckla & Rudel, 1974).

Stimuli. The stimulus set consisted of 36 nonwords written in an artificial script, previously used in Bitan & Booth (2012), in which a pair of symbols represents one letter, and six symbols in different permutations create all six letters of the alphabet: [j-e-l-v-o-p], [artificial script] in the artificial script, respectively. All nonwords were composed of two consonants and one vowel in all possible syllable structures (CVC, VCC, CCV). For example "jop" was written as [artificial script] (see Supplemental Material). Each of the six letters appeared in all possible positions in the words: onset, middle, and final positions. The number of trained words was different for the variable and nonvariable groups. The variable group was trained on 24 nonwords (each letter is repeated 12 times), whereas the two nonvariable groups were trained on only 12 nonwords (each letter is repeated six times). The remaining 12 nonwords were used in the transfer test.

Stimulus presentation and data recording were done using E-Prime software (Psychology Software Tools, 2007).

Procedure

Figure 1 depicts the procedure of the study. Each subject participated in two training sessions, 24 hr apart, including at least 6 hr of sleep during the night. The first session started with exposure to letters, in which the subject was presented with each target letter in the novel script with its corresponding phonological transcription in Latin letters below. Each pair was presented for 2,000 ms, and subjects were instructed to pronounce the related phoneme and memorize the association in order to encourage participants to map the novel script to phonological representations. Each letter was presented once.

This block was followed by exposure to the nonwords. Participants were presented with each target nonword written in the novel orthography for 2,000 ms together with its corresponding phonological transcription in Latin letters below. In order to facilitate mapping of the new words to their phonological representations participants simultaneously heard the correct pronunciation of the nonword through headphones and were required to read it aloud and memorize the association. Each nonword was presented once in the exposure block, resulting in 24 trials in the variable training group. The two nonvariable training conditions (1 and 2) were included in order to equate the exposure of the nonvariable training condition, which included only 12 nonwords. The Nonvariable 1 group received one exposure block (12 trials), thus equating the number of exposures per word to the variable training condition. The Nonvariable 2 group received two exposure blocks (resulting in 24 trials) to equate the total number of trials to the variable training condition.

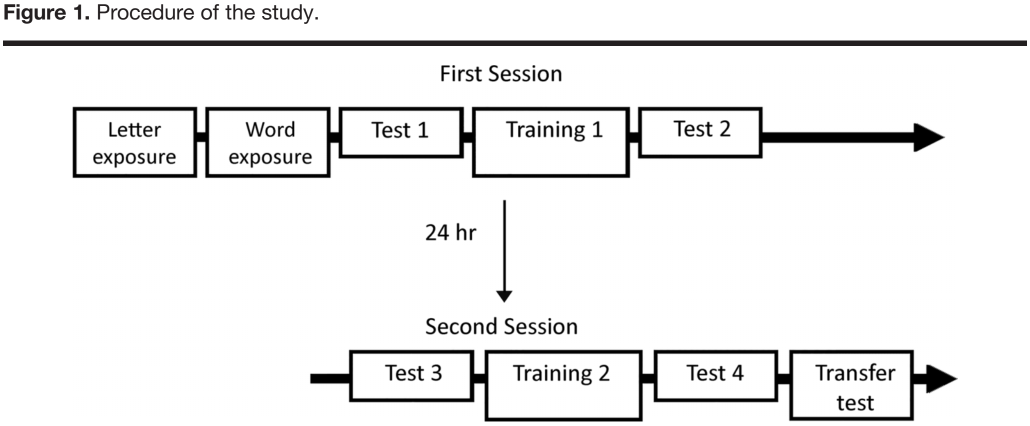
Following the exposure phase, Test 1 was administered. Participants were tested on the nonwords to which they were originally exposed. The test included 48 trials, in which pairings of target nonword and Latin letters appeared for 800 ms. Half of the trials contained correct pairs, and half of the trials were incorrect pairs. The test included one correct trial and one incorrect trial for the variable condition, and two correct and two incorrect trials were implemented in the nonvariable conditions. The subject's task was to judge for each test item whether the Latin-letter string was the correct transcription by pressing one of two buttons on the mouse. This stage lasted about 2 min. No feedback was given for errors.

Test 1 was followed by seven training blocks of 48 trials each, for a total of 336 trials. All training items were presented in a randomized order within each block. In each trial, a target word appeared for 800 ms with a Latin-letter string presented below. Half of the trials contained correct pairs, and half of the trials were incorrect pairs. The task was to indicate whether the pairs were correct by pressing one of two buttons. Auditory feedback was given for errors. Total duration of this phase was about 15 min.

After completing the training, participants performed Test 2, which was identical to Test 1. All tasks were time limited, so if the participant did not respond within 3 s, the trial was recorded as no response.

The second session started with Test 3, followed by seven training blocks, and then by Test 4. At the end of this session, the transfer of learning gains to novel stimuli

Figure 1. Procedure of the study.



was tested. At this phase, subjects were exposed to 12 novel nonwords, presented with the corresponding transcription in Latin script. The exposure block was followed by a test block on untrained nonwords with 48 trials, similar to the trained words tests.

Analysis

In all of our analyses reaction time (RT) for correct responses and accuracy were the dependent measures. Because our main hypothesis was focused on differences between variable and nonvariable training, all of our analyses included planned pairwise comparisons between the variable group and each one of the nonvariable training groups. Therefore, significance was determined based on correction for two comparisons (i.e., $p < .025$). To evaluate performance on the trained items, two general linear model (GLM) analyses (for accuracy and RT) were conducted on the trained items, tests (1–4) as the dependent measures. In addition, to assess differences in the learning curves, we also conducted two GLM analyses (for accuracy and RT) with performance on the training blocks (1–7) as the dependent measure.

In order to test the hypothesis that variable training improves generalization to new words composed of the same letters, we used a one-way analysis of variance (ANOVA) between groups to compare performance on the transfer test as the dependent measure. In addition, we wanted to examine whether the effect of variability on generalization to untrained items can be fully explained by its effect on trained items, or, alternatively, whether variability has a unique advantage for generalization that exceeds its effect on trained items. We therefore calculated a normalized transfer ratio that takes into account the individual's learning gains on trained words (see Bitan & Booth, 2012; Bitan & Karni, 2003, 2004; Bitan et al., 2005). The transfer ratio was calculated for each participant as follows: (performance accuracy in the transfer test – performance accuracy in Test 1) / (performance accuracy in Test 4 – performance accuracy in Test 1). Finally, in order to test the hypothesis that the high variability training group relied more on small unit decoding compared to the nonvariable training groups, we tested effect of phonemic awareness and working memory skills on performance in each group. For this aim we included measures of phonemic awareness and working memory (measured by phoneme deletion time and accuracy and the digit span scores, respectively) separately as covariates in the GLM-repeated measures analyses. This was done (1) for the analyses of accuracy in training blocks and (2) for the analyses of accuracy in the transfer test as the dependent measures.

Results

Performance on Screening Tests

The average performance of the three groups on the screening tests is presented in Table 1. Univariate analysis revealed no main effect of group on the screening tests results: for digit span forward, $F(2, 49) = 0.094$; $p = .91$, for digit span backward, $F(2, 49) = 0.493$; $p = .61$, for reading real words, $F(2, 52) = 0.671$; $p = .51$, for reading nonwords, $F(2, 52) = 0.232$; $p = .79$, for reaction time of phonemic deletion test, $F(2, 52) = 0.532$, $p = .59$, for accuracy of phonemic deletion test, $F(2, 52) = 0.084$; $p = .92$, and for object naming, $F(2, 49) = 0.717$; $p = .49$. These results indicate that all groups have similar basic reading and memory abilities. Planned contrasts comparing the variable training group and each of the two nonvariable training groups revealed no significant differences between groups in any test ($p > .05$).

Trained Items

Accuracy in the Training Blocks

In order to test our hypothesis that variable training would hinder learning of trained words, we conducted a GLM analysis on the accuracy in the training blocks as a dependent variable. Group was the between-subject independent variable and the block number (1–7) and the session number (1–2) were within-subject variables. The analysis showed significant main effects of session, $F(1, 52) = 127.89$; $p < .001$, and of blocks, $F(1, 52) = 50.84$; $p < .001$, and a significant interaction between session and blocks, $F(1, 52) = 31.77$; $p < .001$. The main effect of group across all sessions and blocks was not significant, $F(2, 52) = 2.70$; $p = .07$; planned comparisons between the variable training group and each of the nonvariable groups showed that the variable training group performed more accurately than the Nonvariable 2 group ($p = .03$), but this effect was not significant after correction for multiple comparisons (Figure 2).

To follow up on the interaction between session and block, a separate GLM analysis within each session showed a significant main effect of blocks only in the first session, $F(6, 52) = 21.96$; $p < .001$. Pairwise comparisons showed significant differences between Blocks 1 and 2, $F(1, 52) = 10.93$; $p = .002$; between Blocks 2 and 3, $F(1, 52) = 5.83$; $p = .019$; and between Blocks 4 and 5, $F(1, 52) = 12.58$; $p = .001$. Thus, all groups had improved during the first session, and from the first to the second session.

Accuracy in Trained-Item Tests

In order to continue testing our hypothesis that variable training would hinder learning of trained words, we conducted a GLM analysis with accuracy in the tests as the dependent variable; group as a between-subject variable, and the test number and session number as within-subject factors. Tests 1 and 2 took place before and after training at the first session, Tests 3 and 4 at the second one. The analysis showed significant effects of session, $F(1, 52) = 189.34$; $p < .001$, and test, $F(1, 52) = 237.57$; $p < .001$. The results also showed a significant effect of group, $F(2, 52) = 6.52$; $p = .003$. Furthermore, significant interactions were found between test and session, $F(1, 52) = 34.34$; $p < .001$, and between group and test, $F(2, 52) = 3.26$; $p = .04$. Planned contrasts showed that the performance

Table 1. The average (and standard deviation) of screening tests in each of the three groups and in an independent sample (N = 191) representing local norms.

Test	Group			
	Local norms, N = 191	Variable group, N = 18	Nonvariable 1 group, N = 19	Nonvariable 2 group, N = 18
Digit span forward		10.8 (1.5)	10.7 (1.7)	10.6 (1.9)
Digit span backward		8.5 (1.5)	7.8 (2.5)	8.1 (2.1)
Digit span-combined		19.4 (2.5)	18.5 (3.6)	18.7 (3.4)
Reading real word, no. of correct words in 1 min	106.49 (18.41)	110.8 (17.6)	107.4 (12.2)	113.4 (16.8)
Reading nonreal word, no. of correct nonwords in 1 min	61.04 (14.14)	59.7 (14.0)	61.7 (12.3)	58.8 (13.6)
Phonemic deletion time (s)	109.89 (31.11)	94.8 (20.3)	95.6 (20.3)	89.8 (13.8)
Phonemic deletion, accuracy	22.00 (4.00)	19.8 (4.9)	20.3 (2.1)	19.9 (4.1)
Object naming (s)	33.31 (5.57)	33.4 (5.0)	34.2 (4.6)	32.2 (4.9)

of the variable training group was significantly more accurate compared to the Nonvariable 2 group after correction for multiple comparisons (p = .001), although the comparison to the Nonvariable 1 group (p = .03) did not survive correction for multiple comparisons (Figure 2).

To follow up on the interaction between group and test we conducted a one-way ANOVA within each test, with group as an independent variable. A significant effect for group was found for Test 2, $F(2, 52) = 7.14$; $p = .002$; Test 3, $F(2, 52) = 3.62$; $p = .03$; and Test 4, $F(2, 52) = 6.36$; $p = .003$. The effect of group on Test 1 was marginal, $F(2, 52) = 3.18$; $p = .05$. Planned contrasts showed that the variable training group was significantly more accurate (after correction for multiple comparisons) than the Nonvariable 1 training group in Tests 1 (p = .016) and 2 (p = .008), and from the Nonvariable 2 training group in Tests 2 (p = .001), 3 (p = .011), and 4 (p = .001). As a follow-up analysis on the interaction between test and session, separate analyses within each session showed a significant effect of test in both, the first, $F(1, 52) = 158.88$; $p < .001$, and second, $F(1, 52) = 25.49$; $p < .001$, sessions. Figure 2 shows that the interaction is due to a larger difference between tests in the first session.

Reaction Time in Training Blocks

A GLM analysis, with RT in the training blocks as the dependent variable; group as a between-subject variable and block number (1–7) and session number (1–2) as within-subject variables showed a significant effect for session, $F(1, 52) = 27.82$; $p < .001$, with faster response in Session 2 (see Figure 3). There was no effect for group, $F(2, 52) = 0.029$; $p = .97$, and no significant differences

Figure 2. Accuracy on training blocks and trained item tests for the three groups; open shapes = tests; filled shapes = training blocks. The middle vertical line marks the separation between the two sessions. Error bars indicate standard errors.

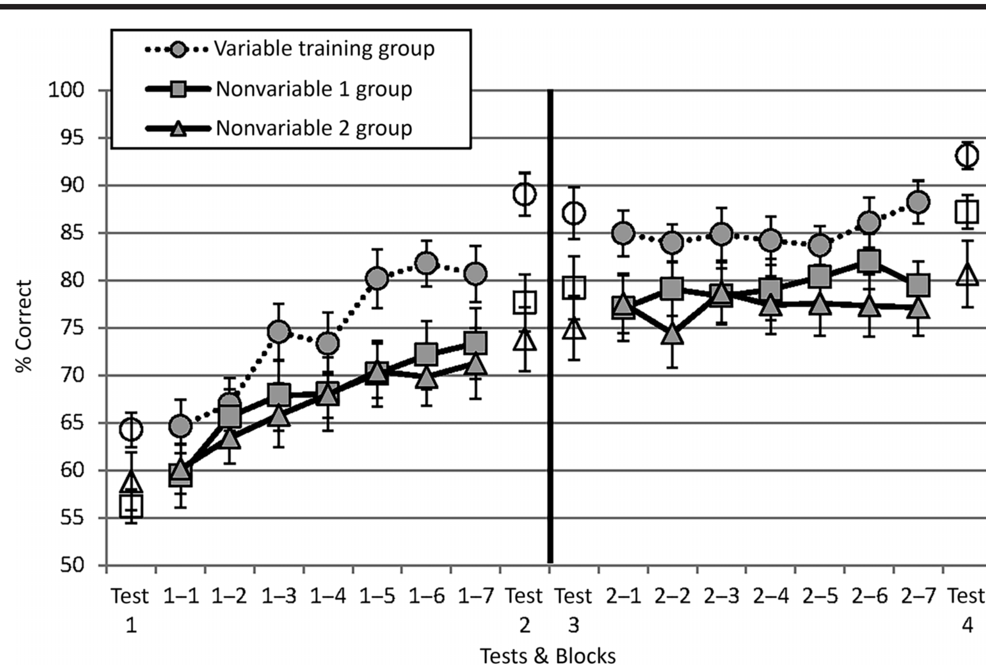
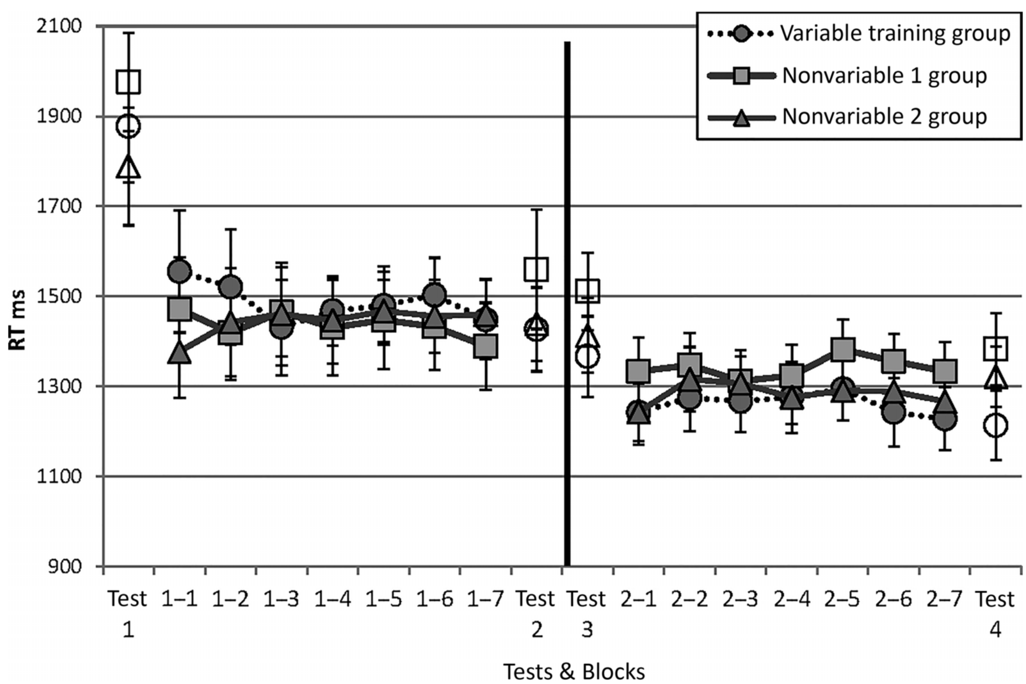


Figure 3. Reaction time (RT) on trained blocks and trained item tests in the three groups; open shapes = tests; filled shapes = training blocks. The middle vertical line marks the separation between the two sessions.



between the variable training group and the other two groups in the planned comparisons.

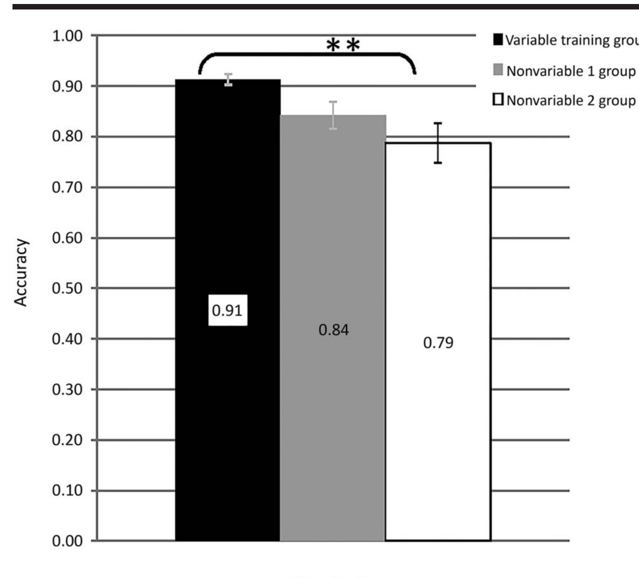
Reaction Time in Trained-Item Tests

A GLM analysis on RT in the tests, with group as a between-subject variable and test and session numbers as within-subject factors showed a significant effect for session, $F(1, 52) = 73.7$; $p < .001$, and test, $F(1, 52) = 48.82$; $p < .001$, and significant interaction between session and test, $F(1, 52) = 15.94$; $p < .001$. There was no effect for group, $F(2, 52) = 0.82$; $p = .44$, and no significant differences between the variable training group and the other two groups in planned comparisons (Figure 3). A separate analysis within each session showed significant differences between tests in both sessions: in the first, $F(1, 52) = 35.99$; $p < .001$, and in the second one, $F(1, 52) = 19.59$; $p < .001$. Figure 3 shows that the effect of test is larger in the first session.

Transfer to New Items

In order to test our hypothesis that variable training would improve generalization to untrained words, a one-way ANOVA was conducted on the accuracy in the transfer test as the dependent variable, with group as a between-subject variable. This analysis showed a significant effect for group, $F(2, 52) = 4.81$; $p = .012$. Planned comparisons showed a significantly higher accuracy for the variable training group compared to the Nonvariable 2 training group after correction for multiple comparisons ($p = .003$) (Figure 4). We also wanted to test the hypothesis that variable training has a unique benefit for generalization to untrained words, beyond its effect on trained words. To test this hypothesis, we calculated a normalized transfer ratio that takes into account the individual's learning gains on

Figure 4. Accuracy of performance in the transfer test. Accuracy values and standard errors are shown. (**) significant difference $p < .01$ corrected.



trained words and compared it between groups. The effect of group was not significant (Figure 5), $F(2, 51) = 2.48$; $p = .78$, and no difference between groups was found in planned contrasts. This indicates that the advantage of the variable training group in performance of the transfer test was due to the higher performance on trained items, and there was no unique benefit for the generalization.

One-way ANOVA conducted on RT during the transfer test showed no significant effect of group on performance, $F(2, 52) = 2.45$; $p = .096$. Planned contrasts showed that the reaction times in the variable training group were shorter than those of the Nonvariable 1 group ($p = .031$), but this did not survive correction for multiple comparisons (Figure 6).

Effects of Working Memory and Phonemic Awareness

Finally, in order to test the hypothesis that the variable training group relies more heavily on decoding of small units, we tested the effect of phonemic awareness and working memory on performance in the different groups. The scores of the digit span and phoneme deletion time and accuracy tests were separately entered as covariates into the GLM repeated-measures analyses. We report here only significant effects of these measures (digit span or phoneme deletion) and their interaction with group.

Working Memory

Including the digit span combined score as a covariate in the analysis of accuracy in training blocks with training block (1–7) as the within-subject factor and group as a between-subject factor showed a significant effect for the digit span scores, $F(1, 46) = 9.56$; $p = .003$, and no interaction with group. Follow-up correlations between digit span and accuracy on training blocks were positive and significant for all blocks (r ranging between .282 and .551 across blocks, $p = .001$ for Block 1; $p = .042$ for Block 2; $p = .002$ for Blocks 3 and 5; $p = .004$ for Blocks 4 and 7; $p < .001$ for Block 6; $p = .009$ for Blocks 8 and 12; $p = .013$ for Block 9; $p = .007$ for Blocks 10 and 11; $p = .043$ for Block 13, and $p = .022$ for Block 14. To test the effect of working memory on generalization, we included the scores of the digit span test as a covariate in the one-way ANOVA for accuracy in the transfer test as the dependent measure. This analysis showed a significant effect for the digit span forward score, $F(1, 46) = 5.66$; $p = .022$, and no interaction with group. A positive correlation was found between digit span forward score and accuracy in the transfer test ($r = .35$, $p = .011$) across groups.

Figure 5. Transfer ratio (normalized to performance on trained items). Accuracy values and standard errors are shown.

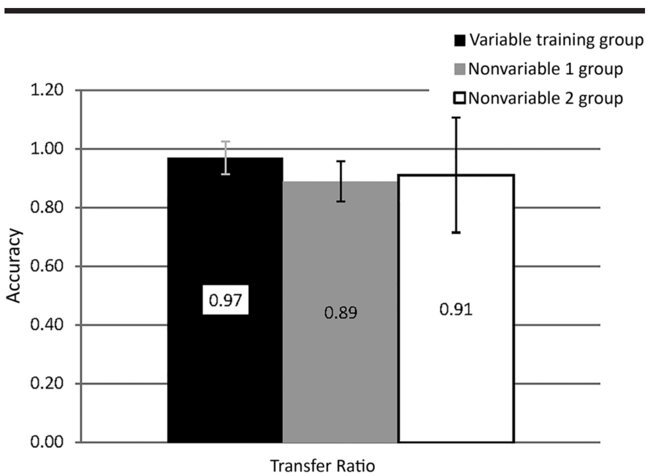
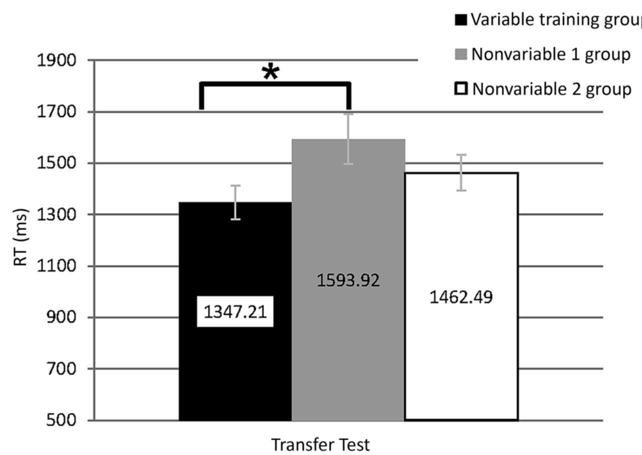


Figure 6. Reaction time (RT) in the transfer test. Mean percent correct and standard errors are shown. (*) significant difference $p < .05$ (uncorrected).



Phonemic Awareness

The scores of the phoneme deletion test (accuracy and time separately) were entered as covariates in the analysis with accuracy in training blocks (1–7) as the dependent measure and group as an independent factor. These analyses showed a significant between-subject effect for time in the phoneme deletion test, $F(1, 49)=7.49$; $p = .009$, as well as for accuracy on this test, $F(1, 49) = 11.67$; $p = .001$, on the accuracy level of the training blocks across groups, and no interaction with group. Correlation analyses indicate that shorter time and higher accuracy in the phoneme deletion test are associated with higher accuracy on training blocks (significant r values range between −0.132 and −0.368, $p < .05$, in 10 out of 14 blocks for time in the phoneme deletion test and between 0.130 and 0.477, $p < .05$, in 10 out of 14 blocks for the accuracy in the phoneme deletion test). To test the effect of phonemic awareness on generalization, we included the scores of the phoneme deletion test as covariates in the analysis of accuracy in the transfer test. This analysis showed a significant effect for accuracy in the phoneme deletion test, $F(1, 49) = 10.97$; $p = .002$, and a significant interaction between this measure and group, $F(2, 49) = 4.13$; $p = .022$. To follow up on this interaction, separate correlation analyses were conducted within each group between accuracy in the phoneme deletion test and accuracy on the transfer test. Positive correlations were found only in the two nonvariable training groups (r = .47, p = .038 for the Nonvariable 1 group and r = .52, p = .025 for the Nonvariable 2 group) but not for the variable training group (r = .031, p = .904).

Discussion

Building on the literature demonstrating a positive effect of variable practice in motor, perceptual, and linguistic learning, we asked how stimulus variability affects learning to read in a novel script. The results of our study confirm the hypothesis that high variability of trained words enhances the process of learning to read and the generalization to reading novel words, suggesting that it promotes the extraction of regularities and the decoding of small orthographic units.

The Effect of Variable Training on Learning to Read Trained Words

The results of our study show that participants in all groups improved during training, achieving 75%–95% accuracy by the end of the second session, indicating they have all learned to read the new words in the novel orthography. Moreover, the benefit of variable training was illustrated by higher accuracy for the variable compared to the nonvariable training groups on the trained-items tests, with no speed–accuracy trade-off.

These findings, indicating the positive effect of variable training even for trained items, are partly inconsistent with our hypothesis and with the results from some of the motor learning (Schmidt & Bjork, 1992) and speech perception (Magnuson & Nusbaum, 2007) studies. These studies show a unique benefit of variable training for generalization to untrained conditions, with a negative effect of variability on performance of trained tasks and on processing of trained stimuli. Other studies showing the advantage of variable training for generalization do not show any effect on performance of trained items when learning morphosyntactic relations in an artificial grammar (von Koss Torkildsen et al., 2013) or when learning a foreign phonemic contrast (Sonu, Kato, Tajima, Akahane-Yamada, & Sagisaka, 2013).

Interestingly, our finding of a positive effect of variability even in trained items is consistent with the findings of Apfelbaum et al. (2013), showing more efficient learning of grapheme-phoneme correspondence of vowels when these are trained in the context of variable consonant frames, for both trained and untrained words. The consistency between our findings and those of Apfelbaum et al. (2013) suggests that the effect of variable training in the reading domain extends also to trained items. Apfelbaum et al. (2013) suggest that the variability of the irrelevant components (different consonants) enhances the readers' focus on the relevant components (vowel letters), thus enhancing their distinctiveness. In the current study, the variable context in which each letter is presented may increase its distinctiveness, thus facilitating the extraction of regularities and increasing the reliance on decoding of letter units. Because decoding of smaller units is an efficient strategy even for reading familiar words, this may explain the benefit of variable training for trained items in the reading domain.

In contrast to the study of Apfelbaum et al. (2013) that manipulated variability in terms of the number of consonants surrounding the target vowels, in the current study the number of letters was fixed across groups and variability was defined by the number of different words in the training set. This enables us to conclude that the number of letter combinations (as manifested in different words), rather than the number of letters, is the critical factor. Moreover, our nonvariable condition included more repetitions on a smaller set of words, which would have an advantage if learning occurred at the level of whole words. Thus, our finding that the benefit of variability overrides the potential benefit of multiple repetitions on whole words further emphasizes the critical importance of small unit decoding even in reading trained words.

It should be noted that the difference in performance on trained items between the variable group and the Nonvariable 1 group was small but significant already on the first test (i.e., before training). This early difference suggests that the benefit of variable practice was already evident after the first block of exposure to the words. In order to control for the possibility that the early advantage of the variable training condition was due to the total number of trials in the exposure block (24 trials in the variable training group and only 12 trials in the Nonvariable 1 group), we included the Nonvariable 2 group, which included two blocks of exposure. However, no difference was found between the two nonvariable training groups. These results suggest that the statistical properties of the stimuli (i.e., its variability) had an effect on the learned units after as little as one exposure.

The Effect of Variable Training on Generalization to Reading Untrained Words

To examine the effect of variable training on generalization, we tested participants' accuracy and reaction time on reading untrained words composed of trained letters at the end of the second session. Consistent with our hypothesis, the variable training group showed the highest accuracy (significant compared to the Nonvariable 2 group) on untrained words, with no speed–accuracy trade-off. In order to test whether there was a unique advantage for variable training on generalization, beyond its effect on learning trained items, we have also computed the transfer ratio, which is an index of generalization normalized to improvement on trained items. The transfer ratio showed no difference between groups, suggesting that there was no unique advantage for variable training on generalization. These results indicate that the advantage of variable training on generalization is fully explained by the same advantage found in trained items. Although this aspect of the results is unexpected, it is due to the unexpected advantage of variable training even for reading trained words as discussed above.

The positive effect of variable training on generalization was previously reported in motor (Douvis, 2005; Wulf & Schmidt, 1997; Yao et al., 2009), perceptual (Amitay et al., 2005; Clopper & Pisoni, 2004; Lively et al., 1993), and linguistic tasks (Apfelbaum et al., 2013; Eidsvåg et al., 2015; Grunow et al., 2006; Perry et al., 2010; Plante et al., 2014; Reeder et al., 2013; Tamminen et al., 2015; von Koss Torkildsen et al., 2013). In the absence of a similar advantage for variability in trained items, this is interpreted as suggesting that variable training reduces the attention allocated to the specific exemplars, thus enabling a more abstract representation of regularities (Perry et al., 2010). This interpretation may also explain the results of the current study, suggesting that high variability induced greater reliance on segmentation and decoding of letters as compared to recognition of whole words. However, because this approach is advantageous also for reading trained words, no unique advantage was found for untrained words. An alternative interpretation for the advantage of variable training for generalization is that the large number of words encountered during training improved participants' ability to learn whole words and learn new information more generally. Whereas the untrained words in the current study were composed of trained letters, this interpretation predicts an advantage of variable training even for untrained words composed of untrained letters, which were not tested in the current study. Nevertheless, the advantage found for variable training for reading trained words cannot be explained by this interpretation.

Phonemic Awareness, Memory, and Generalization

Because we expected participants trained in the variable condition to rely more on decoding of small units and participants trained in the nonvariable condition to rely more on memorization of whole words, we expected differential effects in each group for measures of phonemic awareness and working memory. Including these measures as covariates in the analyses of trained and untrained items revealed two main findings. First, digit span scores, which reflect a general working memory capacity, were positively correlated with performance on trained items across groups. Moreover, digit span forward scores were positively correlated with performance on untrained items across all groups. These results indicate that good working memory contributes to reading acquisition in a novel orthography regardless of variability in the trained stimuli and the applied reading strategy.

The second important point is the effect of phonemic awareness on learning and generalization of the new orthography. Good phonemic awareness was associated with higher accuracy on the trained items across groups as well as better performance on untrained items (transfer test) only for the nonvariable training groups. Phonemic awareness, the ability to segment spoken words into their basic phonemic constituents, is a critical component of reading acquisition, as it underlies the ability to map letters and letter clusters into sounds (Castles & Coltheart, 2004). Although phonemic awareness plays an important role in reading in all orthographies, the correlation between phonemic awareness and successful reading acquisition in children is especially high in deep orthographies, in which the correspondence of letters to sounds is less consistent (Landerl et al., 2013; Ziegler et al., 2010). Thus, although most readers of transparent orthographies can easily map small as well as larger orthographic units to their phonological counterparts, the success of readers of deeper orthographies depends more on having good phonemic awareness. Our results showing a correlation of phonemic awareness with reading of trained items across groups show that this ability still contributes to reading acquisition even in adults learning to read a new orthography (Brennan & Booth, 2015). However, the specific contribution of phonemic awareness to untrained words in the nonvariable training groups is consistent with our prediction that the variable training facilitates the extraction of regularities and decoding of small units. We suggest that participants in the variable training group successfully generalized their decoding knowledge to untrained words regardless of their phonemic awareness level. In contrast, in the nonvariable training groups only those with high phonemic awareness could extract and generalize their knowledge to untrained items.

Conclusion

The results of the current study show that learning a large and more variable set of words improves reading acquisition, compared to receiving more repetitions on a smaller set of words in adults learning to read in a novel orthography. We suggest that this is because the variability reduces the effectiveness of whole word recognition and instead increases the salience of regularities, thus enhancing the extraction of letter-sound correspondences. This effect is consistent with the benefit of variable practice in motor, perceptual, and other cognitive domains and is in line with the idea that distributional properties driving statistical learning are critical factors in learning linguistic categories (Reeder, Newport, & Aslin, 2010; Reeder et al., 2013). However, our results also show domain-specific effects for variable training in reading acquisition, reflected in an advantage for variable training even in reading trained words, with no additional advantage for generalization. This may be due to the critical importance of segmentation, regularities, and small unit decoding for reading.

The practical implications of the current results to reading instruction are clear. Variable training on a large set of words has an advantage over multiple repetitions on a small set of words for adult skilled readers learning to read in a novel orthography. Further studies should test the generalizability of these findings to children learning to read for the first time, but some support can be gained from similar results for children in the studies of Apfelbaum et al. (2013) and Plante et al. (2014). Moreover, our results support the critical importance of small unit decoding in reading acquisition and the effect of phonemic awareness even in skilled adults readers learning to read in a novel orthography. They further demonstrate how a reading instruction method can enhance segmentation and decoding of small orthographic units even when practicing reading of whole words.

Our study has a number of limitations. First, it should be noted that our participants were native Hebrew speakers. Reading instruction of the Hebrew orthography in school uses a very transparent version of the orthography with consistent mapping of graphemes to phonemes. This may have affected the readers' learning strategy and the ease of extracting the correspondence between graphemes and phonemes in the artificial orthography, which would not necessarily generalize to readers whose first language has a nontransparent orthography. A second limitation is the usage of Latin letter transcription of the novel words and letters during training and testing, rather than presenting them only aurally. This was done in order to facilitate learning but may result in mapping the new script only to the familiar letters, instead of mapping them to their corresponding phonological representations. In order to encourage mapping of the new script to phonology, participants have also heard the words and were required to pronounce the phonemes and words during the instruction phase. Our previous studies, using the same paradigm, show that participants' learning of this script is similar to learning a real orthography, as evidenced by correlations with reading scores (Bitan & Booth, 2012) and activation of the reading network in the brain (Bitan et al., 2005). It should also be noted that even if the inclusion of Latin letter transcription facilitated the extraction of regularities and the learning of small units, this effect would apply to all groups and cannot explain the advantage found for variable training. A third limitation is the usage of nonwords with no semantic reference, which may have reduced the reliance on whole word recognition mechanisms. Although this limitation does not undermine the validity of the effect of variable training, it may affect our ability to generalize our findings to real world situations when learning to read a new orthography in a second language.

References

Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Attention, Perception, & Psychophysics, 67,* 691–698.

Apfelbaum, K. S., Hazeltine, E., & McMurray, B. (2013). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology, 49,* 1348–1365. https://doi.org/10.1037/a0029839

Ben Dror, I., & Shani, M. (1996). *Phoneme recognition test for words and pseudowords*. Unpublished test.

Bitan, T., & Booth, J. R. (2012). Offline improvement in learning to read a novel orthography depends on direct letter instruction. *Cognitive Science, 36,* 896–918.

Bitan, T., & Karni, A. (2003). Alphabetical knowledge from whole words training: Effects of explicit instruction and implicit experience on learning script segmentation. *Cognitive Brain Research, 16,* 323–337.

Bitan, T., & Karni, A. (2004). Procedural and declarative knowledge of word recognition and letter decoding in reading an artificial script. *Cognitive Brain Research, 19,* 229–243.

Bitan, T., Manor, D., Morocz, I. A., & Karni, A. (2005). Effects of alphabeticality, practice and type of instruction on reading an artificial script: An fMRI study. *Cognitive Brain Research, 25,* 90–106.

Brennan, C., & Booth, J. R. (2015). Large grain instruction and phonological awareness skill influence rime sensitivity, processing speed, and early decoding skill in adult L2 learners. *Reading and Writing, 28,* 917–938.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91,* 77–111.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech, 47,* 207–238.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach? *Journal of Experimental Child Psychology, 95,* 56–77.

de Graaff, S. D., Bosman, A. M., Hasselman, F., & Verhoeven, L. (2009). Benefits of systematic phonics instruction. *Scientific Studies of Reading, 13,* 318–333.

Denckla, M. B., & Rudel, R. (1974). Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex, 10,* 186–202.

Díaz, G. S., del Rosario Torres, M., Iglesias, J., Mosquera, R., Reigosa, V., Santos, E., . . . Galán, L. (2009). Changes in reading strategies in school-age children. *The Spanish Journal of Psychology, 12,* 441–453.

Douvis, S. J. (2005). Variable practice in learning the forehand drive in tennis. *Perceptual and Motor Skills, 101,* 531–545.

Eidsvåg, S. S., Austad, M., Plante, E., & Asbjørnsen, A. E. (2015). Input variability facilitates unguided subcategory learning in adults. *Journal of Speech, Language, and Hearing Research, 58,* 826–839.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. (2015). The relationship between artificial and second language learning. *Cognitive Science, 40,* 822–847.

Foorman, B. R. (1995). Research on "the great debate": Code-oriented versus whole language approaches to reading instruction. *School Psychology Review, 24,* 376–392.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37–55.

Frost, R. (1992). Orthography and phonology. In P. Downing, S. D. Lima, & M. Noonan (Eds.), *The linguistics of literacy* (pp. 255–274). Philadelphia: John Benjamins Publishing Company.

Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology-Learning Memory and Cognition, 20,* 116–129.

Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders, 39,* 158–170.

Helland, T., Tjus, T., Hovden, M., Ofte, S., & Heimann, M. (2011). Effects of bottom-up and top-down intervention principles in emergent literacy in children at risk of developmental dyslexia: A longitudinal study. *Journal of Learning Disabilities, 44,* 105–122.

Huet, M., Camachon, C., Gray, R., Jacobs, D. M., Missenard, O., & Montagne, G. (2011). The education of attention as explanation of variability of practice effects: Learning the final approach phase in a flight simulator. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1841–1854.

Hulme, C., & Snowling, M. J. (2013). Learning to read: What we know and what we need to understand better. *Child Development Perspectives, 7,* 1–5.

Kyte, C. S., & Johnson, C. J. (2006). The role of phonological recoding in orthographic learning. *Journal of Experimental Child Psychology, 93,* 166–185.

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H., Lohvansuu, K., . . . Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry, 54,* 686–694.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94,* 1242–1255.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 391–409.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 676–684.

Mei, L., Xue, G., Lu, Z., He, Q., Zhang, M., Xue, F., . . . Dong, Q. (2013). Orthographic transparency modulates the functional asymmetry in the fusiform cortex: An artificial language training study. *Brain and Language, 125,* 165–172.

Meschyan, G., & Hernandez, A. (2002). Is native-language decoding skill related to second-language learning? *Journal of Educational Psychology, 94,* 14–22.

Ofen-Noy, N., Dudai, Y., & Karni, A. (2003). Skill learning in mirror reading: How repetition determines acquisition. *Cognitive Brain Research, 17,* 507–521.

Perkins, D. N., & Salomon, G. (1988). Teaching for transfer. *Educational Leadership, 46,* 22–32.

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally. Exemplar variability supports higher-order generalization and word learning. *Psychological Science, 21,* 1894–1902.

Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., . . . Burton, R. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology, 23,* 530–545.

Psychology Software Tools, Inc. (2007). E-Prime (Version 1.2) [Computer software]. Sharpsburg, PA: Author.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2010). Novel words in novel contexts: The role of distributional information in form-class category learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 2063–2068). Austin, TX: Cognitive Science Society.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology, 66,* 30–54.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review, 82*(4), 225–260.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217.

Shapiro, L. R., & Solity, J. (2008). Delivering phonological and phonics training within whole-class teaching. *British Journal of Educational Psychology, 78,* 597–620.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151–218.

Shatil, E. (1995a). *One-minute test for pseudowords*. Unpublished test, Haifa, Israel: University of Haifa.

Shatil, E. (1995b). *One-minute test for words*. Unpublished test, Haifa, Israel: University of Haifa.

Sommers, M. S., & Barcroft, J. (2011). Indexical information, encoding difficulty, and second language vocabulary learning. *Applied Psycholinguistics, 32,* 417–434.

Sonu, M., Kato, H., Tajima, K., Akahane-Yamada, R., & Sagisaka, Y. (2013). Non-native perception and learning of the phonemic length contrast in spoken Japanese: Training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics, 22,* 373–398.

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology, 79,* 1–39.

Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 60–76.

von Koss Torkildsen, J., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *Journal of Speech, Language, and Hearing Research, 56*(2), 618–629.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. New York: Pearson.

Wulf, G., & Schmidt, R. A. (1997). Variability of practice and implicit motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 987–1006.

Yao, W. X., DeSola, W., & Bi, Z. C. (2009). Variable practice versus constant practice in the acquisition of wheelchair propulsive speeds. *Perceptual and Motor Skills, 109,* 133–139.

Yoncheva, Y. N., Blau, V. C., Maurer, U., & McCandliss, B. D. (2010). Attentional focus during learning impacts N170 ERP responses to an artificial script. *Developmental Neuropsychology, 35,* 423–445.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., . . . Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21,* 551–559.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3–29.