Effects of induced orthographic and semantic knowledge on subsequent learning: A test of the partial knowledge hypothesis

Suzanne Adlof[1]
Gwen Frishkoff[2]
Jennifer Dandy[3]
Charles Perfetti[3]

1. University of South Carolina
2. Georgia State University
3. University of Pittsburgh

Published in final edited form: *Reading and Writing*, March 2016, Volume 29, Issue 3, pp 475-500

Published online: January 2016, doi: 10.1007/s11145-015-9612-x

Acknowledgments:

We thank the participants in this study, as well as the teachers and administrators at Falk School at the University of Pittsburgh who assisted with this research. We also thank research assistants Emma Adair, Luree Miller, Leanna Mihalko, Christina Poulin, and Mary Sroka for assistance with data collection, data entry, and reliability checking, and Spencer Babb and Alexis Cattano for their assistance in the preparation of figures. This research was supported by a NIH grant R01HD058566 to the University of Pittsburgh. Additional support was provided during the writing of this manuscript by grants from NIH (R03DC013399) and IES (R305A130467) to the University of South Carolina. The views expressed in this manuscript are those of the authors and do not necessarily represent those of the funding agency.

ABSTRACT

Word learning can build the high-quality word representations that support skilled reading and language comprehension. According to the partial knowledge hypothesis, words that are partially known, a.k.a. "frontier words" (Durso & Shore, 1991), may be good targets for instruction precisely because they are already familiar. However, studies investigating this question have produced mixed findings, and individual differences in baseline knowledge have complicated results both within and across studies. We present two studies that took a different approach, controlling both familiarity and the nature of the familiarizing episode. We controlled familiarity with novel words through pre-exposure ("pre-familiarization") in isolation, to induce form-based familiarity, or in sentences that provided few clues to meaning, to induce partial semantic knowledge. The number of pre-exposures varied (0, 1, or 4). After the pre-familiarization phase, we presented the words in several highly informative sentences to support meaning acquisition. Participants included both adults and typically developing children, ages 9-12. Participants' self-rated familiarity with target words, and their knowledge of the words' meanings and orthography were each measured at baseline, immediately after learning, and one week later. Orthographic and semantic word learning showed contrasting effects of pre-familiarization. For orthographic learning, it was the number, rather than the type, of pre-familiarizations that mattered most. By contrast, the number of pre-familiarizations had little impact on word semantic learning; further, pre-familiarization in low-constraint sentences did not consistently boost subsequent learning. These findings suggest that familiarity with a word prior to instruction does not necessarily improve word-learning outcomes, and they highlight the importance of repeated exposures to high quality contexts for robust word learning.

Key Words: Vocabulary, Word Learning, Partial Word Knowledge, Lexical Quality, Spelling

INTRODUCTION

According to the Lexical Quality Hypothesis (Perfetti & Hart, 2001, 2002), high-quality word representations are characterized by strong, reciprocal links among phonological, orthographic, and semantic knowledge constituents. Therefore, knowledge of one constituent should facilitate learning of another constituent. In vocabulary research, this suggests that partial knowledge of a word might improve subsequent learning of the word's form or meaning (or both). Previous research provides some support for this hypothesis. For example, Rosenthal and Ehri (2009) taught children the pronunciations and meanings of very rare words. Children learned the association of a spoken word form and a picture referent more readily when the picture-word pairs were accompanied by the written word form. Ricketts, Bishop, and Nation (2009) showed similar findings for learning of nonwords using a paired associates paradigm. These studies support the idea that familiarity with a word form can facilitate word learning.

A simple explanation for these findings is that memory representations that engage more links to information in long-term memory are easier to retrieve because there are multiple (phonological, orthographic, or semantic) cues that can activate the representation (Reichle, Rayner, & Pollatsek, 2003). In vocabulary research, a related proposal is that partially known words, a.k.a. "frontier words" (Durso & Shore, 1991), may be good targets for vocabulary instruction precisely because they are already familiar. This view assumes that the cognitive processes in word learning are — in one way or another — different for words that are completely unknown and those that are already associated with a stored memory trace (Reichle & Perfetti, 2003).

Learning of novel versus frontier words

Eye movement studies have shown that readers engage different attentional strategies in learning

words that are already familiar versus those that are completely novel (Chaffin, 1997; Chaffin, Morris, & Seely, 2001; Williams & Morris, 2004). For example, Chaffin et al. (2001) examined eye movements as adult readers were exposed to words that were high familiarity, low familiarity, or novel. The words were embedded in contexts that provided strong cues to word meaning. Readers spent a longer time fixating on and re-reading the novel words as compared with the familiar words. Similarly, Lockett and Shore (2003) found that adults were better able to reject incongruous sentences containing partially known versus completely novel words. In addition, they were more conservative in evaluating correct versus incorrect usage of fully versus partially known words and more conservative in judging known versus novel words.

Frishkoff et al. (2009) observed similar patterns using behavioral and event-related potential (ERP) measures. Participants were to judge whether a particular letter string was or was not a real word. Interestingly, high-skilled readers were more conservative in judging partially known versus novel words. Moreover, their brains responded differently to false rejections versus actual nonwords, suggesting implicit knowledge of some words that they rejected as nonwords (cf. Durso & Shore, 1991). Together, these findings suggest that high-skilled adults are sensitive to the cognitive status of words as fully known, partially known (i.e., familiar), or completely unknown (i.e., novel). In addition, in at least some contexts they use this information to guide reading behavior, either by allocating more attention to items that are less familiar or by adopting a different threshold for identification of somewhat familiar versus novel words.

While there appear to be differences in the processing of frontier vs. novel words (Chaffin et al., 2001; Williams & Morris, 2004), it is less clear that these differences lead to measurable differences in learning. Schwanenflugel, Stahl, and McFalls (1997) failed to observe differences in learning of familiar versus novel words. In their study, 4th-grade children read

several challenging (6th-grade) stories that included words that were known, familiar, or novel based on pretesting of individual participants. Three days later, they completed a multiple-choice test of word knowledge. Scores were higher for both unknown and partially known words that appeared in the story condition versus words that appeared in isolation. In contrast, Jenkins, Stein, & Wysocki (1984) found significant effects of pre-familiarization—in this case, explicit pre-teaching of synonyms for target words—on fifth graders' learning of meanings of unfamiliar words from supportive contexts. Methodological differences between the two studies could help to explain these different outcomes.

One limitation of the studies reviewed thus far is the reliance on pretesting of word knowledge prior to learning. In general, researchers have used variations on the LOWKAT measure developed by Durso & Shore (1991), which provides detailed information about what a person knows about a word, even when she denies that it is part of the lexicon. However, there is also considerable variation in assessment procedures. In addition, there is an inherent trade-off: if a common set of words is used across participants, there are bound to be individual differences in knowledge of some items. On the other hand, selection of different words for each individual based on pretesting can result in word-level variation that is poorly controlled. Given that characteristics of words (concrete/abstract, noun/verb, etc.) can moderate word-learning outcomes, this approach is also less than ideal.

In the present study, we followed the approach of Perfetti, Wlotko & Hart (2005), teaching very rare words that were unlikely to be known by participants. (A slightly different approach is to use pseudowords, e.g., Nation, Angell, & Castles, 2007.) We then presented, or "pre-familiarized," these rare words, prior to learning, either in isolation or in brief written contexts. We also manipulated the number of pre-familiarizations (zero, one, or four). This

allowed us to control the context and amount of pre-familiarization and to thereby circumvent issues due to individual differences in word knowledge at baseline.

The present study

This study has several features that we wish to highlight. First, as mentioned above, we controlled the amount as well as the type of pre-familiarization. Words were pre-familiarized using a form-only (isolated word) exposure or a partial-semantic (low-constraint sentence) exposure. The rationale for low-constraint sentence exposures was that they should represent the quality of incidental exposure to words in real situations, where meaning has not been inferred due either to inattention or lack of context. Thus, although familiarization is a component of word learning, it involves a shallower form-based learning compared with the learning of form-meaning mapping that is the target of word learning. In addition, we varied the number of pre-familiarizations for each condition. This enabled us to look at effects of frequency, as well as type, of pre-familiarization on subsequent word learning. We expected to find superior learning and retention of words that were presented more frequently during the pre-familiarization period, replicating Nation et al., (2007). We also expected to see differences as a function of type of pre-familiarization (words in isolation versus words in context).

Second, we focused on outcomes for contextual word learning (CWL), a complex and poorly understood process that contributes greatly to vocabulary acquisition. CWL is essential for development of educated adult vocabulary — that is, low-frequency abstract words that give us the power to express nuances of meaning and to appreciate a good novel (Beck, McKeown, & Kucan, 2013). Another important feature of CWL is the increased load on meaning comprehension: the focus on meaning may have consequences for acquisition of form (e.g., spelling), particularly for younger or lower-skilled readers. Pre-familiarization could have

different effects for children versus skilled adult readers, who are more fluent and therefore better able to attend to and integrate word- and form-based features, particularly when they encounter a new word in context. Consistent with this idea, Landi, Perfetti, Bolger, Dunlap, and Foorman (2006) found that young readers (children ages 5–8 years) learned and retained new word forms more successfully when the words were trained in isolation rather than in context. The demand on meaning comprehension when "reading" (i.e., pronouncing) words in context may have diverted attentional resources from focusing on the form of the novel word. This finding underscores the importance of age- and skill-related differences in contextual word learning. For this reason, our study participants included adults as well as children (ages 9-12 years).

Our assumption that lexical knowledge builds incrementally in both form and meaning constituents leads to two main hypotheses for this study.

Hypothesis 1: Form-based familiarization. In our study, some words were first presented in isolation. In this case, we expected that the pre-familiarization would yield form-based familiarity, which could be regarded as a shallow form of partial word knowledge (e.g., Chaffin, 1997). Following Ricketts, et al. (2009), we expected that pre-familiarization of the word form would enhance subsequent learning and retention of spelling. Thus, the comparison of interest was the increase in orthographic knowledge from pre- to post-test for trained words that were pre-exposed in isolation (either one or four times), versus trained words that were not pre-familiarized (henceforth, *nonfamiliarized words*).

Hypothesis 2: Partial-semantic familiarization. We also expected different outcomes for words that were pre-familiarized in one or more low-constraint (i.e., weakly informative) contexts versus nonfamiliarized words. In this case, learners may acquire partial knowledge of

the word's meaning, in addition to form-based familiarity (Chaffin, 1997; Frishkoff, Perfetti, & Collins-Thompson, 2010, 2011). Therefore, we predicted that pre-familiarization with words in low-constraint sentences (either one or four times) would lead to better learning and retention of meaning than nonfamiliarized words.

STUDY 1

Method

Participants. Seventy-three children enrolled in grades four through six and 23 adults were recruited from the University of Pittsburgh and its laboratory school and completed at least one session of data collection. Child participants completed sessions with their class as a group, while adults completed sessions individually in the research lab. Due to absences and computer malfunctions, 35 children had incomplete data for one or more sessions. Additionally, 5 adults did not return study after the first study session. Thus, analyses for this study include 38 students enrolled in grades four through six (5 fourth grade, 19 fifth grade, and 14 sixth grade) and 18 adults who completed all sessions and tasks.

Child participants generally exhibited above-average reading comprehension abilities as evidenced by a mean scaled score of 686.14 (SD=33.09) on the *Stanford Achievement Test*-10 Abbreviated Battery Reading Comprehension assessment (Pearson, 2009), with an average percentile rank of 78 relative to national norms, and similarly above-average spelling abilities as evidenced by a mean score of 116.31 (SD=11.41) on the spelling subtest of the *Wide Range Achievement Test-3rd Edition* (WRAT-3; Wilkinson, 1993) and an average percentile rank of 82 relative to national norms. Likewise, adults exhibited strong reading comprehension skills, as evidenced by a mean score of 664.61 (SD=21.40) on the *Gates-MacGinitie-4* Reading

Comprehension assessment (MacGinitie, MacGinitie, Maria, & Dreyer, 2002) and an average percentile rank of 97 relative to national norms for adult community college students.

Child participants completed four sessions, each spaced one week apart. The first three sessions were held in the school's computer lab, where students completed pretests, training and immediate posttests, and delayed posttests for the word learning experiment as a class. The fourth session was held in the regular classroom, where children completed assessments of general spelling and reading comprehension abilities. The procedure for adult participants was very similar, but tasks were administered to individuals, not groups, and the adult assessments of general spelling and reading comprehension abilities typically preceded the word learning sessions.

Word stimuli. Thirty very rare words (10 nouns, 10 verbs, and 10 adjectives), unlikely to have been heard or seen before by elementary aged students, were selected as target words (see Appendix A). These words were a subsample of the target word stimuli in Frishkoff et al. (2010), which had an average frequency of less than one occurrence per million words in both adult and child corpora (Kucera & Francis, 1967; Zeno, Ivens, Millard, & Duvvuri, 1995), and whose meanings could be communicated by a synonym or short phrase that would be familiar to young children. In addition, 15 medium-frequency words were selected as easy filler items for the word knowledge assessments in order to prevent frustration and fatigue (e.g., *evening*, *modern*, *prepare*). These familiar words had an average frequency of 79 and 90 occurrences per million words in child and adult corpora, respectively. As shown in Figure 2, these words were well known by all participants at pre-test, and are henceforth referred to as *known fillers*.

The 30 target words were divided into six lists of five words each. Each word list was then assigned to one of six conditions (see Table 3) in six different versions of the training task

using a Latin square. Each participant experienced each word in only one condition, but across subjects, all words appeared in all conditions.

Sentence stimuli. Four low constraint sentences and two high constraint sentences were selected from Frishkoff, et al. (2010) for each target word (see Table 2 for examples). Low constraint sentences gave very little information about the target word's meaning and were used for sentence familiarization trials (see next section). In contrast, high-constraint sentences provided strong information about word meaning and were used for learning trials. Quantitative measures that capture the level of constraint for each context were based on cloze probability measures (Taylor, 1957) and are reported in Frishkoff, et al. (2010). According to each of these measures, the high-constraint sentences were substantially more constraining — and, hence, more informative — than the low-constraint sentences. Sentences contained approximately 10 words, on average (SD=1.50), and the high and low constraint sentences did not differ in length.

Word Familiarization and Learning Task. The word familiarization and learning task included two types of trials.

On *familiarization trials*, participants viewed target word forms either in isolation (word familiarization conditions) or embedded within a low constraint sentence (sentence familiarization conditions). The number of familiarization trials was also varied (i.e., single exposure conditions vs. 4 exposure conditions).

On *learning trials*, the target word first appeared on the screen within a high-constraint sentence. On the next screen, students were asked to generate a synonym or short phrase to explain the meaning of the word. There were two learning trials for every "trained" word (i.e., for all words except control words).

Familiarization trials and learning trials were randomly mixed throughout the training session; however, all familiarization trials for a given word preceded learning trials for that word. Responses during the learning trials were scored offline using the same partial credit system as for the synonym generation task described below.

Word Knowledge Assessments. Three tasks were constructed to measure children's knowledge of target word forms and meanings at pretest, immediate posttest and delayed posttest. These assessments were always administered in the order described. All tasks were computer administered at all test times, except for the spelling pretest.

Spelling. At pretest, the examiner read each word aloud, and students were instructed to spell the word on their answer sheet. At the immediate and delayed posttest, the spelling task was administered over the computer using E-prime 2.0 stimulus presentation software (Schneider, Eshman, & Zuccolotto, 2007). Digital recordings of the target words featured the examiner's voice, and students were allowed ten seconds to write each word on their answer sheet. Words appeared in a fixed random order at each test point. Responses were scored offline by trained research assistants. Each test was scored by two assistants, with scoring discrepancies resolved by discussion with the first author.

Familiarity Rating and Synonym Generation. On each trial, a target word appeared at the center of the screen, and students were asked to rate their familiarity with the target word, on a scale of zero (e.g., "I don't know this word, and I've never seen it before") to nine (e.g., "I know this word, and I've seen it many times"). After giving a familiarity rating, students were instructed to "*Type one word that means the same thing as the target word. If you can't think of just one word, you may write a short phrase.*" Word presentation order was randomly determined by the stimulus presentation software, and students had 10 seconds per word to give their

familiarity ratings and 30 seconds per word to generate a synonym. Participant responses were scored on a partial credit scale by trained research assistants. Fully correct answers featuring a clear synonym or specific definition for the target word received three points; incomplete or nonspecific answers demonstrating medium knowledge for the word received two points; antonyms or responses that suggested knowledge of the general domain of the word received one point; nonresponses and fully incorrect responses received zero points. For example, for the target word *blench*, a response of "dodge" would receive three points, "move" would receive two points, and "gasp" would receive one point. Responses were double scored by two trained research assistants to ensure reliability; disagreements were resolved by discussion with the first author.

Synonym Matching. In this task, students were asked to select from five choices the word or phrase closest in meaning to the target word. All correct answers and foils contained high-frequency words known to children. For each item, there was at least one phonological/orthographic foil and one meaning-related foil. The synonym-matching task was computer-administered at all three test times, and words were presented in random order by the stimulus presentation software. Responses were scored by the E-Prime program.

Data Analysis. Child and adult data were analyzed separately. To examine whether pre-familiarization influenced performance during the learning trials, within-subjects analysis of variance was used to compare the five conditions of trained words (recall that there were no learning trials for control words). To examine changes in word knowledge (i.e., learning and/or forgetting) across test times, we conducted a 3 (Time) by 6 (Condition) repeated-measures analysis of variance for each assessment. Significant main effects and interactions were followed up using paired comparisons or simple effects analysis, as described in the Results section.

Results demonstrating significant increases in scores across test sessions would provide evidence for learning, whereas results demonstrating significant differences between words in the control condition relative to the other five conditions would provide evidence that the learning was caused by the training itself. Additionally, results indicating significant differences in favor of one of the pre-familiarization conditions over the non-familiarized condition would provide evidence of a benefit of induced familiarity.

Results

Learning Trials. On each learning trial, participants read a high constraint sentence and then were asked to generate a synonym for the target word. Figure 1 displays synonym generation scores by training condition for both children and adults (see also Supplemental Tables 1-2). Results indicated no significant differences between training conditions in synonym generations during learning trials for children [F (4, 148) =0.53, p = .72, η_p^2 =.01] or for adults [F (4, 68) = 0.78., p = .55, η_p^2 =.04]. Thus, there appeared to be no significant advantage for having prior exposure to words in isolation or in low constraint sentences prior to completing the learning trials.

Word Knowledge Assessments. Mean scores for all word knowledge assessments by Time and Condition are displayed for child and adult participants in Figure 2. Mean and standard error values for all data points are also provided in Supplemental Tables 3-4.

Spelling. Mean spelling accuracy scores are displayed in the top row of Figure 2. Analysis of child scores indicated a significant effect of Time [F (2, 74) = 153.97, p <. 001, η_p^2 =.81]. There was no main effect of Condition [F (5, 185) = 1.30, p =. 27, η_p^2 =.03] or interaction between Condition and Time [F (10, 370) = 1.42, p =. 17, η_p^2 =.04]. Pairwise comparisons of each test time revealed that scores at pretest were significantly lower than on the immediate post-

test (p<.001) and were also lower than on the delayed posttest (p<.001). Furthermore, scores at the delayed posttest were significantly higher than those at the immediate posttest (p=.003). Overall, these results suggest that children's spelling accuracy for all words improved across test times, but the lack of differences between trained words and control words in the overall analysis was unexpected. Visual inspection of Figure 2 suggests that children's accuracy of spelling words in the control condition was lower than for other conditions at immediate posttest but not pretest or delayed post-test. Follow-up comparisons at immediate posttest confirm this finding: spelling accuracy of words in the control condition was significantly lower than the 1-sentence, 1-word, and 4-sentence condition ($p \leq .01$); all other comparisons were non-significant (all p >.07).

Results for adults indicated a significant main effect of Time [F (2, 34) = 144.95, p <. 001, η_p^2 =.90] and a non-significant main effect of Condition [F (5, 85) = 1.83, p =. 12, η_p^2 =.10]. These effects were qualified by a significant interaction between Time and Condition [F (10, 170) = 2.02, p =. 03, η_p^2 =.11]. Pairwise comparisons of performance at each test time indicated that spelling improved significantly from pretest to immediate posttest (p <.001), and scores remained stable between immediate posttest and delayed posttest (p = .14). To examine the Time by Condition interaction, we examined Condition effects separately within each test time. There were no significant Condition differences at pretest [F (5, 85) = .46, p =. 81, η_p^2 =.03] or delayed posttest [F (5, 85) = 1.84, p =. 19, η_p^2 =.10]. However, significant differences were observed at the immediate posttest [F (5, 85) = 5.27 p <. 001, η_p^2 =.24], when spelling accuracy for control words was significantly lower than all of the other five conditions (all p <.02) which did not significantly differ from each other (all p > .11).

Taken together, these results indicate that both children and adults showed improvements in spelling accuracy across all words—including control words—between the pre-test and the two posttests. Although words in the control condition (which only appeared on tests, and never appeared in the training program) were spelled significantly less accurately than words in the other five conditions at immediate posttest, by delayed posttest, they were equivalent to scores in the other conditions. Finally, there was no difference in accuracy for any of the trained word conditions. Words that had been pre-familiarized, whether in isolation or in low-constraint sentences, were no more likely to be accurately spelled than words in the non-familiarized conditions.

Perceived Familiarity. Average familiarity ratings by Time and Condition are plotted in the second row of Figure 2. Statistical analysis of child ratings revealed a significant main effect of Time [F (2, 74) = 6.80 p =. 002, η_p^2 =.16]. There was no main effect of Condition [F (5, 185) = 1.32, p =. 26, η_p^2 =.03]. The effect of Time was qualified by a Time by Condition interaction [F (10, 370) = 3.46, p <. 001, η_p^2 =.09]. Follow-up analyses showed no Condition differences at pretest [F (5, 185) = .83, p =.53, η_p^2 =.02]. Condition effects were significant at the immediate posttest [F (5, 185) = 4.66, p<.001, η_p^2 =.11]. In particular, control words were rated as significantly less familiar than trained words (all $p \leq .02$). However, these differences in perceived familiarity were no longer significant by the delayed posttest [F (5, 185) = 1.08, p =.37, η_p^2 =.03]. The different types of trained words did not differ from each other (all $p > .14$).

Adults showed a similar pattern, with significant main effects of Time [F (2, 34) = 82.28, p <. 001, η_p^2 =.84] and Condition [F (5, 85) = 5.28, p <. 001, η_p^2 =.25], qualified by a significant interaction [F (10, 170) = 9.40, p <. 001, η_p^2 =.37]. At pre-test, there were no significant Condition differences in ratings of familiarity [F (5, 85) = .16, p =.98, η_p^2 =.01]. Significant

Condition differences were observed at immediate posttest [F (5, 85) = 14.87, p <. 001, η_p^2 =.47], when control words were rated as less familiar than words in all other conditions (p <.001), and words in the 4-sentence condition were rated as more familiar than words in the 1-word condition (p =.01) and the 4-word condition (p =.03). No other Condition differences reached significance (p > .05). There was a trend for words in the 4-sentence condition to be rated higher than words in the non-familiarized and 1-sentence conditions (both p = .06). By delayed posttest, these differences were no longer significant [F (5, 85) = 2.15, p =.07, η_p^2 =.11].

Taken together, these results suggest that perceptions of familiarity may be transient and more reflective of perceived semantic knowledge than the number of prior exposures. Following training, children and adults perceived trained words as being significantly more familiar than control words. However, within the set of trained words, there were no significant differences in children's perceived familiarity for words based on pre-exposure status. Though some small differences in perceived familiarity were detected in adults at immediate posttest, these were non-significant at the delayed posttest.

Synonym Generation. Descriptive data for the synonym generation task are plotted in the third row of Figure 2. As expected, child and adult ratings at pretest were essentially at floor for unknown words across conditions. Both groups showed improvement following training, with adults showing greater improvement than children. Analysis of child data indicated significant effects of Time [F (2, 74) = 41.93 p <. 001, η_p^2 = .53] and Condition [F (5, 185) =6.48, p <. 001, η_p^2 =.15], which were both qualified by a significant interaction [F (10, 370) = 5.19, p < .001, η_p^2 =.12]. There was no Condition effect at pre-test [F (5, 185) =1.83, p = .11, η_p^2 =.05]. Significant Condition differences were detected at the immediate posttest [F (5, 185) =6.48, p <. 001, η_p^2 =.15], when performance in the control condition was lower than all other conditions (all p

<.001), which were not different from each other (all p > .15). Significant Condition differences were also detected at the delayed posttest [F (5, 185) =3.81, p =. 003, η_p^2 =.09], when words in the control condition were significantly worse than the 4-sentence condition (p<.001) and all other conditions (p < .03). There was also a small advantage for words in the 4-sentence condition relative to the non-familiarized (p = .01) and 4-word (p = .03) conditions. All other comparisons between training conditions at the delayed posttest were non-significant (p > .07).

Adult scores showed main effects of Time [F (2, 34) = 76.03, p <. 001, η_p^2 =.83] and Condition [F (5, 85) = 12.79, p <. 001, η_p^2 =.44], which were qualified by a significant interaction between Time and Condition [F (10, 170) = 9.03, p <.001, η_p^2 =.36]. Following up the interaction, there were no significant Condition differences at pretest [F (5, 85) = .24, p =. 95, η_p^2 =.02], but significant effects were detected at immediate posttest [F (5, 85) = 15.78, p <. 001, η_p^2 =.48] and delayed posttest [F (5, 85) = 5.42, p <. 001, η_p^2 =.24]. At the immediate posttest, scores for control words were significantly lower than for all other conditions (p < .001). Additionally, scores for the 1-sentence condition were significantly higher than the 4-word condition (p = .02), and all other comparisons were non-significant (all p > .08). At the delayed posttest, scores for control words were significantly lower than words in all conditions (all $p \leq$ .01). Additionally, scores for the non-familiarized condition were significantly higher than scores for the 4-word condition (p =.02). All other comparisons were non-significant (all p > .06).

In summary, these results indicate that children and adults were able to learn and retain information about word meanings after encountering words in high-constraint contexts. In the child sample, there was a small benefit of prior exposure to words in the 4-sentence condition at delayed posttest; whereas for adults, there was a small disadvantage for words pre-exposed in the

4-word condition. Thus, the effects of prior exposure conditions on synonym generation performance showed some variation across participant groups and across time.

Synonym Matching. The final row of Figure 2 displays child and adult mean accuracy scores by Condition and Time for the synonym matching task. Child results indicated significant main effects of Time [F (2, 74) = 16.76 p <. 001, η_p^2 =.31] and Condition [F (5, 185) = 4.81, p <. 001, η_p^2 =.12], which were both qualified by a significant interaction [F (10, 370) = 2.51, p = .006, η_p^2 =.06]. To explore this interaction, Condition effects were tested separately at each Time. There were no Condition differences at the pre-test [F (5, 185) = .44, p =.82, η_p^2 =.01], but significant Condition differences were detected at both immediate [F (5, 185) = 6.08, p <. 001, η_p^2 =.14] and delayed posttest [F (5, 185) = 3.00, p =. 013, η_p^2 =.08]. At the immediate posttest, control words were less accurate than words in all other conditions (all p <.005). There was also a small advantage for words in the 4-sentence condition compared to the 1-word condition (p = .029). All other comparisons between conditions at immediate posttest were non-significant (p > .05). At the delayed posttest, words in the control condition were significantly less accurate than words in the non-familiarized (p = .015), 1-word (p = .023), and 4-sentence (p =.004) conditions, but not the 1-sentence (p=.189) or 4-words (p = .220) conditions. No other comparisons between conditions reached significance (all p > .06).

Adult scores showed main effects of Time [F (2, 34) = 73.33, p <. 001, η_p^2 =.81] and Condition [F (5, 85) = 9.50, p <. 001, η_p^2 =.36]. These effects were qualified by an interaction between Time and Condition [F (10, 170) = 8.29, p <.001, η_p^2 =.33]. To examine the interaction, we examined Condition effects within each Time separately. There were no significant differences between conditions at pretest [F (5, 85) = .46, p =. 81, η_p^2 =.03]. However, differences between conditions were observed at the immediate posttest [F (5, 85) = 14.19, p <.

001, η_p^2 =.46] and delayed posttest [F (5, 85) = .95, p <.001. η_p^2 =.36]. At both posttests, words in the control condition were significantly less accurate than words in all other trained conditions (all $p < .001$), but accuracy of words in the other trained conditions did not differ from each other (all $p > .16$).

Together, these results converge with those of the synonym generation task in showing that both children and adults learned and retained word meanings after exposure to high constraint sentence contexts. Evidence for an effect of pre-familiarization was weaker than for the synonym generation task. Children continued to display a slight advantage at immediate posttest for words in the 4-sentences pre-familiarization condition, but it was only significant relative to words in the 1-word condition. Adults showed no significant effects of pre-familiarization.

Interim Summary and Discussion

The purpose of Study 1 was to examine the effects of induced familiarity with words prior to contextual learning of form and meaning. Results indicated that both children and adults were able to learn and retain word meanings after two exposures to high constraint sentences. That is, there were substantial gains in knowledge for trained words versus untrained words. However, pre-familiarization had little effect on self-rated familiarity with words or on acquisition of meaning. In the child sample, there appeared to be a small advantage for learning the meanings of words that had been pre-exposed in four low constraint sentences, as compared with words that had not been pre-exposed, but this advantage was not observed in the adult group. Instead, the adult participants showed a slight disadvantage for words appearing in the 4-word condition, relative to the non-familiarized condition in the delayed post-test synonym generation task. One possibility is that adults experienced interference from the familiarization

contexts at the delayed post-test point.

Spelling outcomes showed a different pattern of results. Recall that words assigned to the control condition appeared on pre- and post-tests, but did not appear in the familiarization or learning phases. Not surprisingly, there was no change in performance on these words across test times on the familiarity ratings, synonym generation, or synonym matching tasks. However, both children and adults showed growth in spelling performance for control words across test times. These results indicate that exposure to the control words during the pre- and post-test assessments provided adequate exposure to increase knowledge of orthography, suggesting that relatively little exposure is sufficient to build orthographic form knowledge in older children and adults. Taken together, these results suggest that different types of exposure are important for acquisition of word forms vs. word meanings. The acquisition of word meanings required exposure to high constraint sentences. There was no evidence that pre-familiarization with word forms in isolation benefitted the acquisition of word meanings, although there was some evidence that sentence familiarization helped children learn and retain word meanings. In contrast, the acquisition of form (spelling) simply required exposure to the form constituents, and the type of pre-familiarization had little influence.

Considering these results, we questioned whether the design of the familiarization task had encouraged participants to pay more attention to sentence familiarization trials than to word familiarization trials. If so, this could have obscured differences in learning due to pre-familiarization. While familiarization trials occurred before learning trials for each target word, the familiarization and learning trials were intermixed across words. Moreover, during the learning trials, participants were asked to generate a synonym for the target word following the presentation of the word in a high constraint sentence. However, there was no signal to indicate

whether an upcoming trial was a familiarization or learning trial. Thus, learning trials were more similar to sentence familiarization trials than to word familiarization trials. Participants may have learned that a sentence could be followed by a request for synonym generation, whereas an isolated word would never be followed by such a request. This could have motivated participants to attend more to the sentence familiarization trials than to the word familiarization trials, which would explain the advantage for the 4-sentence familiarization condition for children and the disadvantage for the 4-word condition for adults. To address this possible confound, we conducted a follow-up study with children that included active processing tasks for word and sentence familiarization trials (see Study 2).

STUDY 2

The purpose of Study 2 was to replicate and extend the findings of Study 1 with a new sample of children using the same basic design, with minor modifications. Our primary goal in Study 2 was to determine whether pre-familiarization with a word in four low constraint sentences would benefit subsequent learning when word and sentence familiarization trials were equally likely to be followed by a prompt for active retrieval of the word. To address this question, we made three methodological changes. First, we attempted to boost power for detecting condition differences by dropping the 1-word and 1-sentence familiarization conditions, which were of lesser interest, given the lack of differences between these conditions and the nonfamiliarized condition in Study 1. Second, we attempted to increase children's learning of the trained words by decreasing the number of words to be learned from 25 to 16 and by increasing the number of learning trials from two to three. Last, we separated familiarization trials from learning trials and included active tasks throughout to encourage attention to both the word and sentence familiarization trials.

Method

Participants. Participants included 20 children enrolled in fourth (n=11) and fifth (n=9) grades at the university laboratory school who provided complete data across all training and testing sessions. An additional 28 children completed at least one study session, but were excluded from the data analysis due to missing data related to absences or computer malfunction. On average, child participants exhibited above average spelling abilities as evidenced by mean scores of 110.05 (SD = 16.58) on the spelling subtest of the WRAT-3 (Wilkinson, 1993), and typical vocabulary knowledge as evidenced by mean standard scores of 11.95 (SD = 4.24) and 11.50 (SD = 3.82) on the receptive and expressive vocabulary subtests of the *Test of Word Knowledge* (Wiig & Secord, 1990). As in Study 1, participants completed four study sessions, spaced at one week intervals, in which pretests, training and immediate posttests, delayed posttests, and assessments of general spelling and reading comprehension abilities, respectively, were completed.

Word and Sentence Stimuli. Word stimuli included 16 rare word stimuli (8 nouns, 8 verbs) also used in Study 1 (see Appendix A). Four additional rare words (2 nouns, 2 verbs) from Study 1 were included as filler words for the active familiarization tasks. Finally, ten familiar words from Study 1 (5 nouns, 5 verbs) were included as easy fillers for the word knowledge assessments to minimize frustration and fatigue.

The 16 target words were divided into four lists of four words each. Each word list was then assigned to one of four conditions (see Table 3) in four different versions of the training task using a Latin square. Each participant experienced each word in only one condition, but across subjects, all words appeared in all conditions. Similar to Study 1, four low constraint sentences and three high constraint sentences were selected for each target word.

Word Familiarization and Learning Task. In Study 2, the familiarization trials were completed for all words before the learning trials. Each familiarized target word appeared four times, either in isolation or in a low constraint sentence. In addition, rare filler words and familiar filler words were presented in active familiarization trials, to encourage attention to the words and sentences. On active sentence familiarization trials, filler words were first presented in a low constraint sentence; then on the next screen, participants were asked to generate a synonym for the target word they had just seen. On active word familiarization trials, filler words were first presented in isolation; then on the next screen, students were asked to retype the word that they had just seen. Familiarization trials for target words and filler words were presented in a random order determined by the stimulus presentation software.

Word Knowledge Assessments. Children's word knowledge was assessed one week prior to training, immediately following training, and one week after training using the same three assessments as in Study 1, i.e., spelling, familiarity rating, synonym generation, and synonym matching. One small modification to the spelling assessment was made: In Study 2, all three spelling assessments were computer administered. Digital recordings of the target words featured the examiner's voice, and students were given ten seconds to write each word on their answer sheet. Words appeared in a fixed random order at each test point. At pretest, participants were given the option to replay the recording of the target word a second time; however, they were given only one presentation opportunity at all subsequent test times.

Results

Learning trials. Children's performance by training condition is displayed in Figure 3. Within-subjects ANOVAs revealed no significant differences between training conditions in children's synonym generations during learning trials [F (2, 38) = 0.24, p =.79, η_p^2 =.01]. These

results replicated the finding of Study 1, in that there appeared to be no advantage for pre-familiarization to words in isolation or in low constraint sentences prior to completing the learning trials.

Spelling. Children's mean spelling accuracy scores by Condition and Time are displayed in Figure 4. Results indicated a significant effect of Time [F (2, 38) = 35.94, p <. 001, η_p^2 =.65]. There was no significant effect of Condition [F (3, 57) = 2.22, p =. 10, η_p^2 =.11], but the Time by Condition interaction was significant [F (6, 114) = 2.91, p =. 01, η_p^2 =.13]. To explore the interaction, Condition differences were examined separately at each test time. The effect of Condition was non-significant at pretest [F (3, 57) = 0.96, p =. 34, η_p^2 =.05] and delayed posttest [F (3, 57) = 1.98, p =. 13, η_p^2 =.09], but it was significant at intermediate posttest [F (3, 57) = 12.96, p <. 01, η_p^2 =.40], when control words were spelled less accurately than words in all other conditions (all p < .02), which did not differ from each other (all p >.41). As in Study 1, these results suggest that there was significant growth in accuracy of spelling words assigned to the control condition between pre- and post-test. Follow-up *t*-tests confirmed that control words were spelled more accurately at delayed post-test than at both immediate posttest (p =.02) and pretest (p =.01), which did not differ from each other (p =.11).

Familiarity Rating. Children's mean ratings of word familiarity by Time and Condition are plotted in Figure 4. As in Study 1, the target words received relatively low familiarity ratings at all test times. Analysis revealed significant main effects of Time [F (2, 38) = 8.15, p <. 001, η_p^2 =.30] and Condition [F (3, 57) = 7.32, p <. 001, η_p^2 =.28]. The Time by Condition interaction did not reach statistical significance [F (6, 114) = 1.96, p =.08, η_p^2 =.09]. Pairwise comparisons indicated that the main effect of Time was due to significantly lower familiarity ratings at pretest than at both immediate (p <.001) and delayed posttest (p = .01). The main effect of Condition

was due to significantly higher familiarity ratings for the 4-word condition than for all other conditions (all $p \leq .05$) and significantly lower familiarity ratings for the control condition than for the 4-word ($p < .001$) and nonfamiliarized ($p = .01$) conditions. Although the interaction was not statistically significant, post hoc *t*-tests confirmed that there were no significant differences between conditions in familiarity ratings at pre-test (all $p > .07$). The finding of higher familiarity ratings for the 4-word condition than all other conditions was different from Study 1 where control words were rated as less familiar than all other words, suggesting that words presented in isolation, in the presence of a task drawing attention to forms may have boosted perceived familiarity, though further study is required for confirmation. However, the finding of lower familiarity ratings for control words is consistent with Study 1.

Synonym Generation. Children's mean synonym generation scores by Time and Condition are plotted in Figure 4. As in Study 1, performance was at floor at pretest, and scores improved but remained low across immediate and delayed posttests. Results indicated significant effects of Time [F (2, 38) = 35.94, $p < .001$, $\eta_p^2 = .65$] and Condition [F (3, 57) = 6.12, $p < .001$, $\eta_p^2 = .24$], which were qualified by a significant interaction [F (6, 114) = 3.33, $p < .01$, $\eta_p^2 = .15$]. There were no significant Condition differences at pretest [F (3, 57) = 1.05, $p = .38$, $\eta_p^2 = .05$]. However, significant differences were detected at immediate post test [F (3, 57) = 4.57, $p < .01$, $\eta_p^2 = .19$], when performance for words in the 4-word condition was significantly better than for all other conditions (all $p < .03$), which did not differ from each other (all $p > .15$), and at delayed post test [F (3, 57) = 4.99, $p < .01$, $\eta_p^2 = .21$], when performance for control words was significantly worse than all other conditions (all $p < .03$), which did not differ from each other (all $p > .20$). While the pattern of results at delayed posttest was largely consistent with findings from Study 1, the finding of better performance on words in the 4-word condition at immediate

posttest was not. Perhaps drawing attention to word forms using an active task temporarily boosted recall for those words; however, this benefit was not retained over time.

Synonym Matching. Children's mean synonym matching scores by Time and Condition are plotted in Figure 4. Results indicated significant effects of Time [F (2, 38) = 20.11, p <. 001, η_p^2 =.51] and cCndition [F (3, 57) = 8.11, p <. 001, η_p^2 =.30], which were qualified by a significant Time by Condition interaction [F (6, 114) = 3.42, p <. 01, η_p^2 =.15]. There were no significant Condition differences at pretest [F (3, 57) = .69, p = .56, η_p^2 =.04], but significant differences were detected at immediate posttest [F (3, 57) = 16.92, p <. 001, η_p^2 =.47], when accuracy for words in the control condition was significantly lower than for all other conditions (all p <.001), which did not differ from each other (all p > .83). These Condition differences were no longer statistically significant at the delayed posttest [F (3, 57) = 2.54, p =. 07, η_p^2 =.11]. These findings were largely consistent with those of Study 1, in that there were large effects of training, but no significant differences for prior familiarization.

GENERAL DISCUSSION

The goal of this work was to determine whether prior familiarity with words facilitates contextual learning of spelling or meaning, or both. This research draws on theories of partial word knowledge (Durso & Shore, 1991) and on the Lexical Quality framework of Perfetti and Hart (2001, 2002), which suggest that partial word knowledge could facilitate contextual word learning. Previous studies investigating this question have produced mixed findings, perhaps due to heterogeneity in the level of knowledge individuals had for stimulus words at the study outset. Thus, we used extremely rare words and confirmed through pretesting that learners had no prior knowledge of these words. Next, we induced familiarity with some words by first presenting them, either in isolation, or in one or four low-constraint sentences, which provided few clues to

meaning. Finally, we presented trained words in either two or three high-constraint sentences, to support contextual learning of meaning, as well as form.

Acquisition of meaning from high-constraint contexts

An important outcome of this study was that participants — both children and adults — readily learned the meanings of words from only two or three high constraint sentences followed by an active synonym-generation task. In Study 1, synonym-generation scores for trained words (averaging across conditions) increased from pre- to post-test for adults (Figure 2, row 3, column 2), and for children (Figure 2, row 3, column 1). By comparison, synonym-generation scores for control words remained at floor. Synonym matching showed the same pattern (Figure 2, row 4). In addition, participants retained a portion of this new knowledge over the one-week period that separated the immediate and delayed post-test. Similar patterns of learning and retention were observed in Study 2 (Figure 4).

These findings replicate prior work that demonstrates efficient and effective learning of word meanings from single-sentence contexts, particularly when learners are highly engaged and when the contexts are sufficiently directive to support lexical inferencing (Frishkoff, et al., 2008, 2011). Further, they contrast with studies of "incidental" CWL, where students are typically engaged in text comprehension rather than lexical inferencing and where the text itself provides unreliable or too few cues to support robust learning (Beck, et al., 2013).

On the other hand, there were notable differences in the extent of word semantic learning for the children versus the adults. First, children derived less information about word meanings than adults from high constraint sentences, as indicated by their synonym generations during learning trials. Second, children showed lower performance than adults across posttests that assessed meaning knowledge. In Study 1, on the easier (synonym-matching) test, adults were just

above baseline (25% accuracy) at pretest and showed an increase to around 73% accuracy on the immediate post-test. By contrast, synonym-matching scores for children increased substantially less, to around 36%. On the harder (synonym-generation) test, scoring of partial semantic knowledge suggested that both groups were substantially below ceiling on the immediate post-test. In Study 2, when children had fewer words to learn (16 vs. 25) and an extra learning opportunity for each trained word (3 vs. 2 learning opportunities), posttest performance was much improved, although still lower than that for adults.

Across both age groups, the general implication is that learning from up to three high-constraint exposures remains incomplete, with a form-meaning connection that is not yet robust. While this partial learning may facilitate comprehension of new texts containing these words, additional informative exposures may be needed to establish robust representations of word meanings, i.e. sufficient to be expressed in productive meaning assessments such as meaning generation.

Metacognitive biases in familiarity ratings

Although familiarity ratings are not our central focus for this study, these data can help inform our understanding of the other outcome measures because they involve self-ratings of knowledge at pre- and post-test and therefore reflect metacognitive states (i.e., feeling-of-knowing), as well as actual learning. In Study 1, children's familiarity ratings for trained words were higher than those for control words at immediate posttest, but there was no significant difference between any of the pre-familiarized conditions vs. the non-familiarized condition; moreover, differences between trained words and control words were no longer statistically significant at the delayed posttest. Adults also showed differences in self-rated familiarity for trained versus control words at immediate posttest, and showed a non-significant trend for greater perceived familiarity in the

4-sentence condition than the non-familiarized condition. However, contrary to our expectations, these differences were not retained at delayed posttest. It is possible that both child and adult participants were conservative in their self-ratings of familiarity. They required a relatively high level of confidence in their semantic knowledge before giving a high rating, consistent with the assumption that meaning is more salient than form for native speakers of a language (Marcel, 1983). Thus, participants may implicitly assign more weight to their knowledge of a word's meaning (particularly knowledge that is accessible to short-term memory), and give less consideration to form-based familiarity.

In Study 2, words in the control condition continued to be rated as less familiar than words in all other conditions, but children reported higher ratings of perceived familiarity for trained words in the 4-word condition than for trained words in other conditions. These findings suggest that when children's attention was drawn to word forms due to the presence of an active processing task for filler word items, their perceived familiarity ratings may have reflected form-based familiarity. Thus, the estimates from Study 2 may give a more accurate estimate of changes in familiarity from pre- to post-test.

Effects of pre-familiarization on spelling learning

Beyond the overall pattern of learning from high-constraint contexts, we predicted that familiarity with a word's form—that is, pre-exposure to the words in isolation— would increase knowledge of a word's orthography. This hypothesis was not supported: There was no difference detected between pre-familiarized and non-familiarized words at any test time. In fact, as illustrated in Figure 4, children and adults showed increased knowledge of spelling from pre- to post-test for all words, including those in the control condition. In addition, although there was a small advantage for trained versus control words on the immediate post-test, it was no longer

evident on the delayed post-test. Note that control words appeared only during testing. Thus, at the immediate post-test, control words had been seen twice in the previous week, whereas all other words had been seen between four (in the nonfamiliarized condition) and eight (in the 4-sentence and 4-word conditions) times. The increase in orthographic knowledge from such few exposures (six total exposures over three weeks for control words) is striking, suggesting that acquisition of specific word forms is rapid for readers with sufficient experience in reading and thus a well developed orthographic lexicon. The other implication is that the type of pre-familiarization is not all that important for orthographic learning. Instead, it seems to be the number of exposures that matters most.

Effects of pre-familiarization on meaning learning

The pattern of results was strikingly different for acquisition of word meanings. In Study 1, children enjoyed a slight advantage in learning the meanings of words that were pre-familiarized in four low-constraint sentences vs. nonfamiliarized words. Interestingly adults showed a slight disadvantage for words that were pre-familiarized four times in isolation vs. non-familiarized words. It is possible that adults tried harder to make sense of the familiarization contexts and subsequently experienced interference; however, this will need to be explored in future studies. Because the familiarization and learning task design in Study 1 may have implicitly encouraged attention to sentences and inattention to isolated words, we conducted Study 2, which included active processing tasks for filler words in isolation and in low constraint sentences. Children in Study 2 showed an advantage for words pre-familiarized in the 4-word condition; however, this advantage was not maintained at the delayed posttest. Overall, even when the effects of pre-familiarization were significant, they were subtle relative to the effects of training in high constraint sentences in both studies.

These findings suggest that low-quality exposures are not sufficient to boost acquisition of the word's meaning. Perhaps this seems obvious, since the contexts that we used in the pre-exposure phase provided minimal clues to meaning. On the other hand, it has relevance for theories of partial word knowledge, which have predicted differences in learning of frontier versus novel words. Although our low-constraint contexts induced partial knowledge of the sort that Durso and Shore (1991) have emphasized, they did not lead to improved CWL, suggesting that, while individuals may possess varying degrees of knowledge for any given word, having a shallow level of knowledge does not automatically facilitate the acquisition of fully specified semantic representations. Our confidence in this conclusion, however, can be only as strong as the measures we used to assess the acquisition of word meanings. We included both an easier receptive task (synonym matching) and a harder generative task (synonym generation) to measure knowledge of word meanings. It is possible that other measures of semantic knowledge would have produced a different pattern of results. However, across the two studies and two age groups, there was little evidence that pre-familiarization improved performance on either task.

Finally, we consider the relation between the kind of learning required by our experimental task and everyday outside-the-laboratory word learning. Both everyday and academic learning of new words often entail new concepts as well as new forms. Our experimental task appears to encourage attaching new words to existing concepts. For example a learner can infer from a high constraint context that "purloin" means the same as "steal": *Someone broke into our house and **purloined** all our valuables.* As an initial learning event, certainly forming a meaning equivalence between "purloin" and "steal" is sufficient. With further experience with contexts using "purloin" the learner may refine its sense within the broad meaning of "steal" to refer more specifically to stealing with stealth. Such learning, which is

captured by the episodic word learning model of Reichle and Perfetti (2003), is very common. Children and adults come to know words that refine (add or subtract relevant features) of more familiar words; e.g., learning adjectives for that refine positive emotions beyond the general word "happy" to "content," "elated," or "euphoric" (e.g., Beck, McKeown, & Kucan, 2013). Thus the kind of learning represented in this study is a typical form of word learning through verbal context in its early phases. Whether the limitations we observe in the role of word form familiarization would also apply to new conceptual learning, which often happens with direct instruction rather than incidental learning with verbal contexts, remains to be seen.

Conclusion

The goal of word learning is to build and retain high-quality lexical representations, which can then support skilled reading and language comprehension. The present study considers both the quality and quantity of exposures to a word, prior to instruction. The study is unique in that is the first to parametrically manipulate "partial knowledge," thus providing a more fine-grain test of the partial knowledge hypothesis, compared with previous studies.

Our findings suggest that quality and quantity of exposure play different roles in the acquisition of lexical form and meaning. For orthographic learning, it is the number of exposures, rather than the type of exposure, that matters. By contrast, the quality (i.e., informativeness) of a context is a major factor in word semantic learning. Even a single high-constraint context can support acquisition of meaning: In fact, so-called "fast-mapping" of form to meaning is common in early childhood, when novel word forms are readily mapped to concrete referents. For higher-level abstract words, however, multiple contexts are needed to promote robust knowledge of meaning. Thus, our results are consistent with the view that high quality contexts are integral to robust acquisition of meaning.

REFERENCES

Beck, I., McKeown, M., & Kucan, L. (2013). *Bringing Words to Life: Robust Vocabulary Instruction*. 2nd ed. New York: Guilford Press.

Chaffin, R. (1997). Associations to unfamiliar words: Learning the meanings of new words. *Memory & Cognition, 25*(2), 203-226.

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: a study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 225.

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Experimental Psychology: General, 120*(2), 190.

Frishkoff, G. A., Collins-Thompson, K., Perfetti, C. A., & Callan, J. (2008). Measuring incremental changes in word knowledge: Experimental validation and implications for learning and assessment. *Behavioral Research Methods, 40*(4), 907-925.

Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2010). Lexical quality in the brain: ERP evidence for robust word learning from context. *Developmental Neuropsychology, 35*(4), 376-403.

Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2011). Predicting Robust Vocabulary Growth from Measures of Incremental Learning. *Scientific Studies of Reading, 15*(1), 71-91.

Frishkoff, G. A., Perfetti, C. A., & Westbury, C. (2009). ERP measures of partial semantic knowledge: left temporal indices of skill differences and lexical quality. *Biological Psychology, 80*(1), 130-147.

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal, 21*(4), 767-787.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Landi, N., Perfetti, C, Bolger, D., Dunlap, S., & Foorman, B. (2006). The role of discourse context in developing word form representations: A paradoxical relation between reading and learning. *Journal of Experimental Child Psychology, 94*(2), 114-133.

Lockett, J. N., & Shore, W. J. (2003). A narwhal is an animal: Partial word knowledge biases adults' decisions. *Journal of Psycholinguistic Research, 32*(4), 477-496.

MacGinitie, W.H., MacGinitie, R.K., Maria, K., & Dreyer, L.G. (2006). *Gates-MacGinitie Reading Tests, 4th Edition.* Itaska, IL: Riverside.

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15*(2), 197-237.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning words from context during normal reading. *American Educational Research Journal, 24*(2), 237-270.

Nation, K., Nation, P., Angell, A., & Castles. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology, 96*(1), 71-84.

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfien (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67–86). Washington, D. C.: American Psychological Association.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of Functional Literacy* (Vol. 11, pp. 67–86). Amsterdam: John Benjamins.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*(6), 1281-1292.

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word-experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading, 7*(3), 219-237.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *Behavioral & Brain Sciences, 26*(4), 445-476; discussion 477-526.

Ricketts, J., Bishop, D. V., & Nation, K. (2009). Orthographic facilitation in oral vocabulary acquisition. *The Quarterly Journal of Experimental Psychology, 62*(10), 1948-1966.

Rosenthal, J., & Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology, 100*(1), 175.

Schneider, W., Eschman, A., & Zuccoloto, A. (2007). *E-prime 2.0.* Psychological Software, Inc.

Schwanenflugel, P. J., Stahl, S. A., & Mcfalls, E. L. (1997). Partial word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research, 29*(4), 531-553.

Stein, J. M., & Shore, W. J. (2012). What do we know when we claim to know nothing? Partial knowledge of word meanings may be ontological, but not hierarchical. *Language and Cognition, 4*(3), 141-166.

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental Word Learning While Reading: A Meta-Analysis. *Review of Educational Research, 69*(3), 261-285. doi: 10.3102/00346543069003261

Taylor, W. L. (1953). "Cloze" procedure: a new tool for measuring readability. *Journalism Quarterly, 30*, 415.

Wagovich, S.A., & Newhoff, M. (2004). The single exposure: Partial word knowledge growth through reading. *American Journal of Speech-Language Pathology, 13*, 316-328.

Wiig, E. H. & Secord, W. (1990). *Test of Word Knowledge*. San Antonio, TX: Pearson.

Wilkinson, G. S. (1993). *Wide range achievement test (WRAT3).* Washington, DE: Wide Range Inc.

Williams, R., & Morris, R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology, 16*(1-2), 312-339.

Zeno, S., Ivens, S., Millard, R., & Duvvuri, R. (1995). The educator's word frequency guide. Touchstone Applied Science Associates (TASA).

Table 1. Description of word learning conditions in Study 1

Condition Name	Number of Familiarization Trials	Type of Familiarization Trials	Number of Learning Trials
Control	0	-	0
Non-familiarized	0	-	2
1-word	1	Word	2
4-word	4	Word	2
1-sentence	1	Sentence	2
4-sentence	4	Sentence	2

Table 2. Sample word and sentence stimuli.

Target Word	Low Constraint Sentences (Familiarization Trials)	High Constraint Sentences (Learning Trials)
Jactancy (noun)	***Jactancy*** *is something you can see in people of all ages.* *It's easy to see when someone else is showing* ***jactancy****.* *People noticed that Josey tended to display* ***jactancy*** *at odd times.* *Sally tried very hard to hide her* ***jactancy*** *with her family.*	*People who only talk about themselves are full of* ***jactancy****.* *It's too bad the winner showed* ***jactancy*** *instead of humility.*
Purloin (verb)	*She told her friend that she had* ***purloined*** *when she was a kid.* *The man in the tan jacket* ***purloined*** *the paperwork before he left.* *We tried to figure out whether the object had been* ***purloined****.* *Bob did not think it would be necessary to* ***purloin*** *it.*	*Someone broke into our house and* ***purloined*** *all our valuables.* *The child was punished for trying to* ***purloin*** *candy from the store.*
Nocent (adjective)	*Tom often takes up activities that some people think are* ***nocent****.* *The flower that we picked had a lot of* ***nocent*** *qualities.* *We had no idea whether the material was* ***nocent*** *or not.* *It is not hard to find* ***nocent*** *objects in this collection.*	*Clearing broken glass can be* ***nocent*** *if you aren't careful.* *Many foods were* ***nocent*** *to him because of his allergies.*

Table 3. Description of word learning conditions in Study 2

Condition Name	Number of Familiarization Trials	Type of Familiarization Trials	Number of Learning Trials
Control	0	-	0
Non-familiarized	0	-	3
4-word	4	Word	3
4-sentence	4	Sentence	3

Figure 1. Mean synonym generation scores during learning trials by children and adults in Study 1.

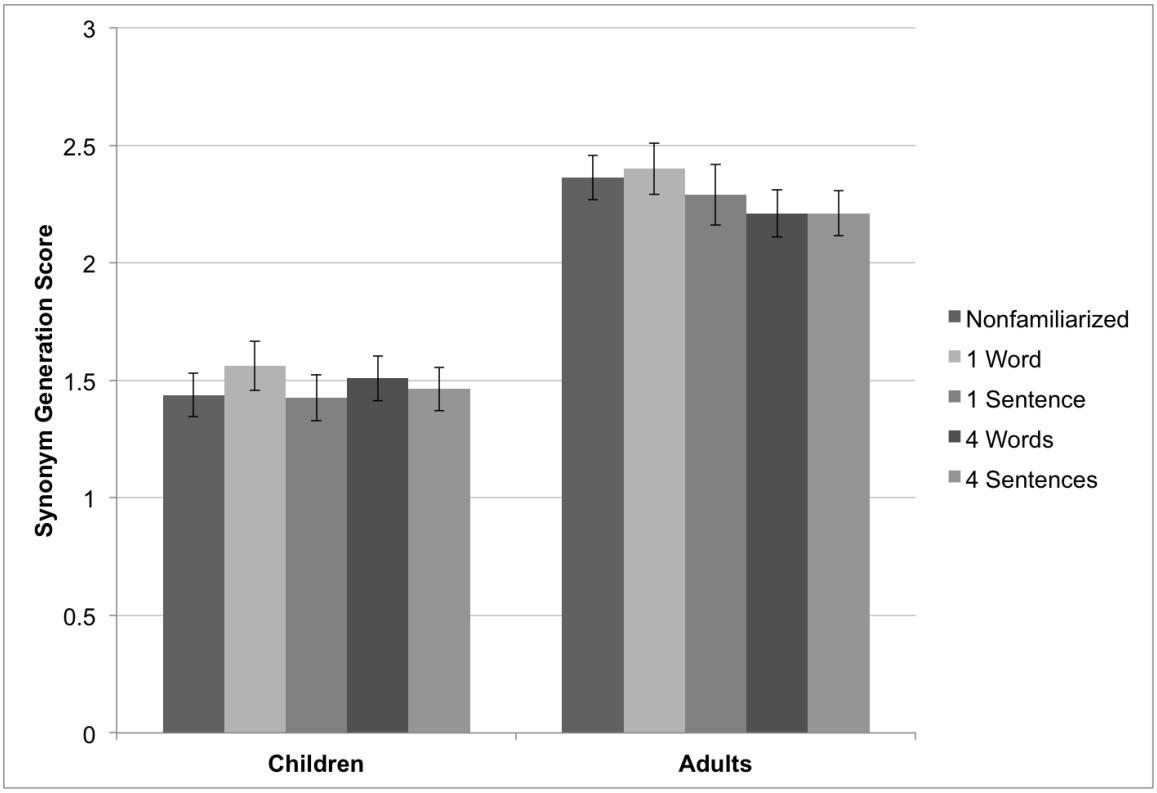


Figure 2. Mean word knowledge scores for adults and children in Study 1

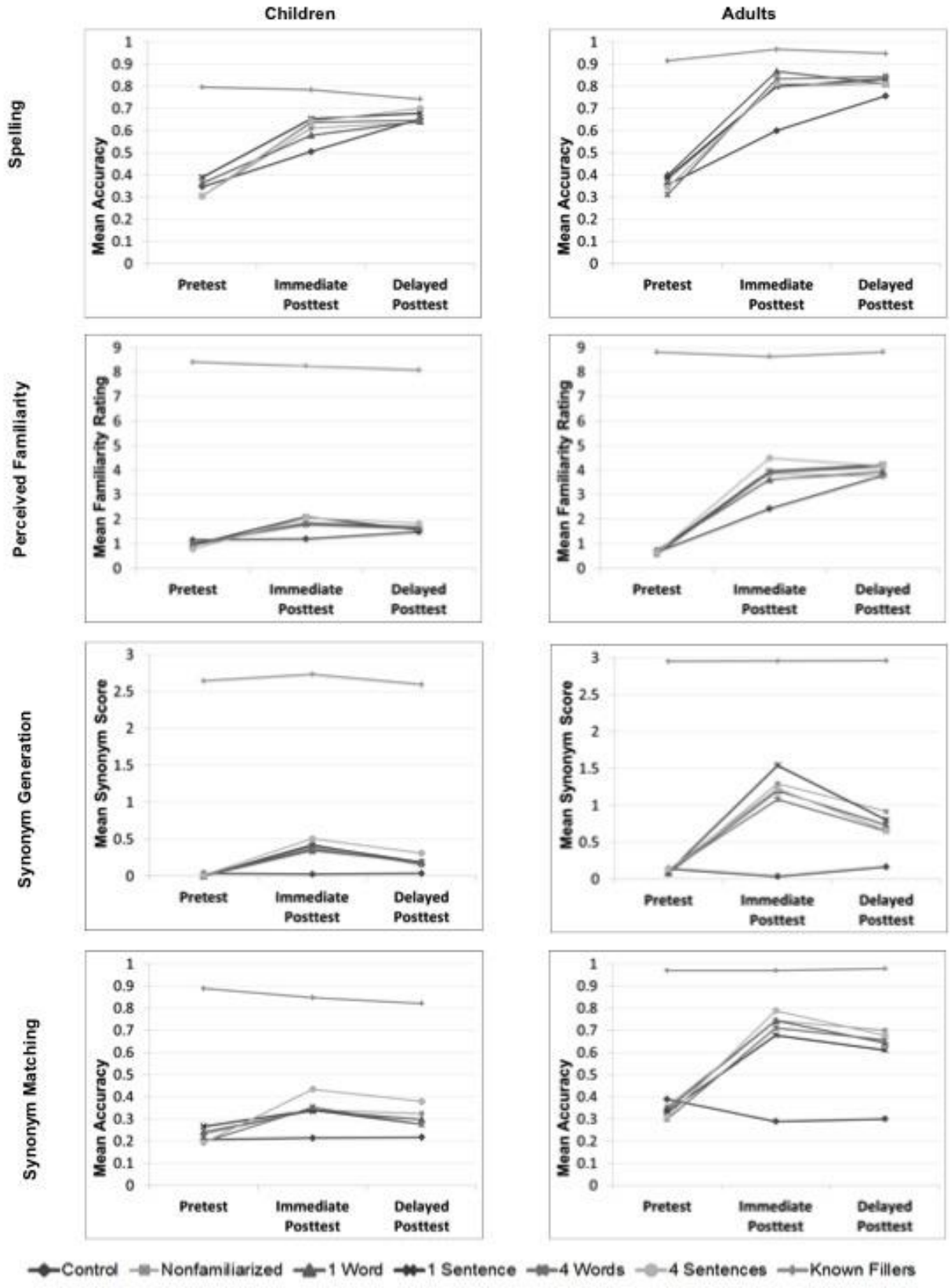


Figure 3. Mean synonym generation scores during learning trials by children in Study 2

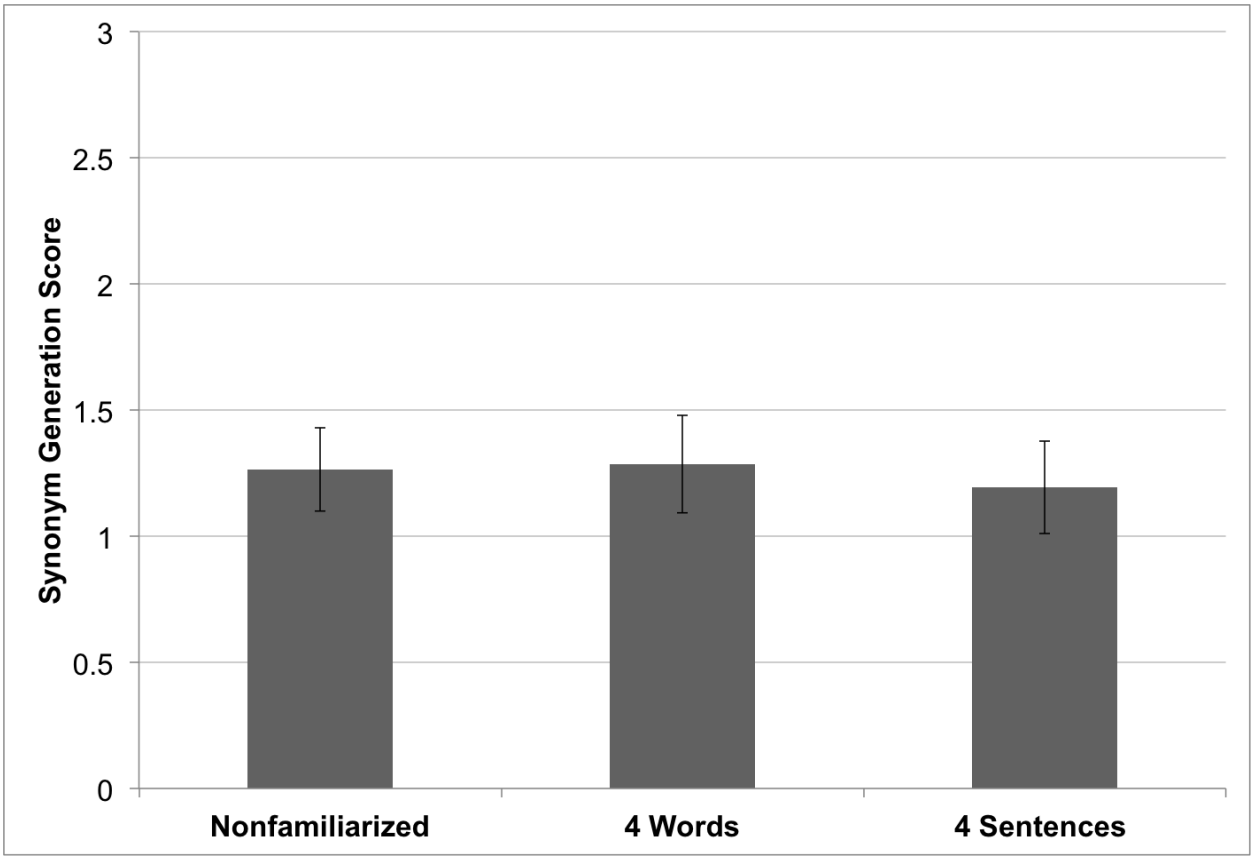
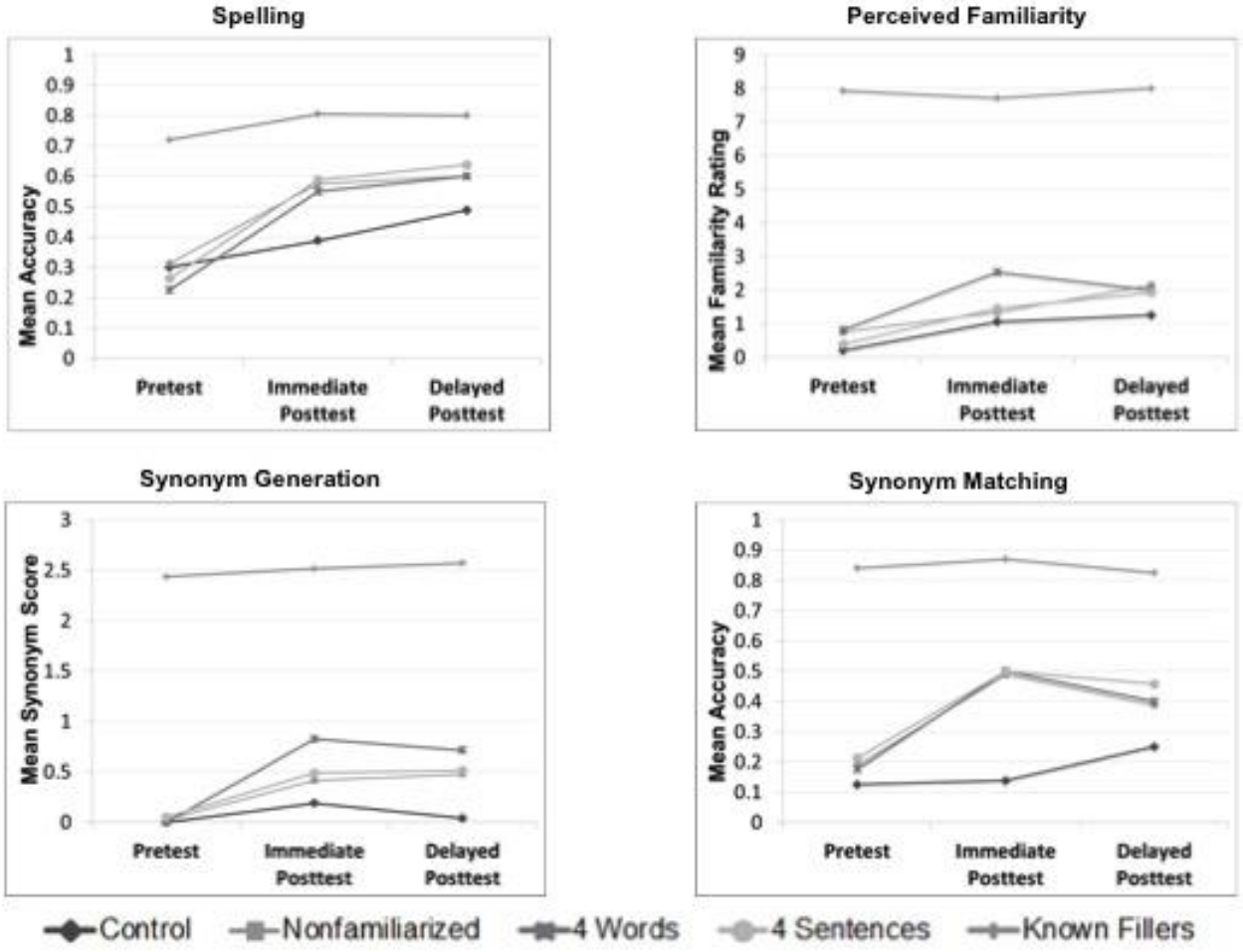


Figure 4. Mean word knowledge scores for children in Study 2



Appendix A. Stimulus Words for Study 1

Stimulus Word	Part of Speech
blench+	verb
burke+	verb
chouse*	verb
conticent	adjective
debouch*	verb
dehort*	verb
enation*	noun
esculent	adjective
fingent	adjective
fulgor	noun
glozing*	verb
gramercy+	noun
impavid	adjective
jactancy*	noun
kippage*	noun
legerity*	noun
lenitive	adjective
macilent	adjective
mundify*	verb
nocent	adjective
nutant	adjective
plangent	adjective
priscan	adjective
proditor+	noun
purloin*	verb
repine*	verb
susurrus*	noun
swink*	verb
viridity*	noun
wanion*	noun

* Words included as target words in Study 2.
+ Words included as hard filler words in Study 2.

Supplemental Tables

Supplemental tables provide means and standard errors for each data point included in Figures 1-4.

S.1. Learning Trial Scores for Children in Study 1 (Max = 3)

Condition	Mean	SEM
Nonfamiliarized	1.44	0.09
1 Word	1.56	0.11
1 Sentence	1.43	0.10
4 Words	1.51	0.10
4 Sentences	1.46	0.09

S.2. Learning Trial Scores for Adults in Study 1 (Max = 3)

Condition	Mean	SEM
Nonfamiliarized	2.36	0.09
1 Word	2.40	0.11
1 Sentence	2.29	0.13
4 Words	2.21	0.10
4 Sentences	2.21	0.10

S.3. Word Knowledge Assessment Performance for Children in Study 1

		Pretest		Immediate Posttest		Delayed Posttest	
Assessment	Condition	Mean	SEM	Mean	SEM	Mean	SEM
Spelling (Proportion Correct)	Control	0.35	0.04	0.51	0.04	0.65	0.04
	Nonfamiliarized	0.31	0.04	0.61	0.04	0.64	0.04
	1 Word	0.36	0.04	0.58	0.04	0.64	0.04
	1 Sentence	0.39	0.04	0.65	0.04	0.68	0.04
	4 Words	0.31	0.03	0.64	0.04	0.65	0.04
	4 Sentences	0.31	0.04	0.64	0.04	0.70	0.04
	Known Fillers	0.80	0.02	0.78	0.02	0.74	0.03
Perceived Familiarity Rating (Max = 9)	Control	1.15	0.30	1.19	0.22	1.48	0.34
	Nonfamiliarized	1.03	0.24	1.86	0.27	1.67	0.38
	1 Word	0.96	0.12	1.83	0.28	1.67	0.34
	1 Sentence	0.97	0.23	2.08	0.31	1.55	0.33
	4 Words	1.04	0.20	1.77	0.28	1.61	0.30
	4 Sentences	0.79	0.17	2.05	0.26	1.83	0.34
	Known Fillers	8.42	0.24	8.24	0.25	8.08	0.26
Synonym Generation (Max = 3)	Control	0.04	0.02	0.03	0.02	0.04	0.02
	Nonfamiliarized	0.01	0.01	0.43	0.09	0.15	0.04
	1 Word	0.00	0.00	0.35	0.07	0.19	0.06
	1 Sentence	0.01	0.01	0.41	0.09	0.18	0.06
	4 Words	0.00	0.00	0.37	0.07	0.17	0.05
	4 Sentences	0.01	0.01	0.51	0.10	0.31	0.06
	Known Fillers	2.64	0.05	2.73	0.03	2.60	0.10
Synonym Matching (Proportion Correct)	Control	0.21	0.02	0.21	0.02	0.22	0.02
	Nonfamiliarized	0.23	0.02	0.34	0.03	0.32	0.03
	1 Word	0.24	0.03	0.34	0.03	0.30	0.03
	1 Sentence	0.27	0.03	0.34	0.03	0.28	0.03
	4 Words	0.20	0.02	0.35	0.03	0.28	0.03
	4 Sentences	0.19	0.02	0.43	0.04	0.38	0.03
	Known Fillers	0.89	0.02	0.85	0.03	0.82	0.03

S.4. Word Knowledge Assessment Performance for Adults in Study 1

		Pretest		Immediate Posttest		Delayed Posttest	
Assessment	Condition	Mean	SEM	Mean	SEM	Mean	SEM
Spelling (Proportion Correct)	Control	0.36	0.05	0.60	0.05	0.76	0.04
	Nonfamiliarized	0.38	0.05	0.80	0.04	0.81	0.04
	1 Word	0.40	0.04	0.87	0.04	0.81	0.06
	1 Sentence	0.39	0.06	0.80	0.04	0.83	0.05
	4 Words	0.31	0.06	0.83	0.05	0.84	0.04
	4 Sentences	0.34	0.05	0.81	0.04	0.81	0.04
	Known Fillers	0.92	0.04	0.97	0.01	0.95	0.02
Perceived Familiarity (Max = 9)	Control	0.66	0.16	2.42	0.30	3.77	0.38
	Nonfamiliarized	0.60	0.17	3.86	0.30	3.77	0.38
	1 Word	0.73	0.13	3.60	0.27	3.95	0.37
	1 Sentence	0.67	0.17	3.96	0.33	4.22	0.32
	4 Words	0.57	0.22	3.92	0.34	4.16	0.41
	4 Sentences	0.65	0.20	4.49	0.40	4.18	0.40
	Known Fillers	8.82	0.09	8.64	0.20	8.82	0.09
Synonym Generation (Max = 3)	Control	0.14	0.05	0.04	0.03	0.17	0.06
	Nonfamiliarized	0.08	0.05	1.29	0.20	0.92	0.16
	1 Word	0.08	0.04	1.20	0.14	0.73	0.11
	1 Sentence	0.09	0.06	1.54	0.20	0.80	0.12
	4 Words	0.12	0.06	1.08	0.18	0.65	0.16
	4 Sentences	0.13	0.06	1.22	0.16	0.66	0.13
	Known Fillers	2.95	0.02	2.95	0.02	2.96	0.02
Synonym Matching (Proportion Correct)	Control	0.39	0.04	0.29	0.05	0.30	0.05
	Nonfamiliarized	0.36	0.04	0.74	0.06	0.70	0.05
	1 Word	0.34	0.05	0.74	0.05	0.64	0.06
	1 Sentence	0.33	0.05	0.68	0.06	0.61	0.06
	4 Words	0.30	0.05	0.71	0.06	0.66	0.06
	4 Sentences	0.31	0.04	0.79	0.07	0.68	0.04
	Known Fillers	0.97	0.01	0.97	0.01	0.98	0.01

S.5. Learning Trial Scores for Children in Study 2 (Max = 3)

Condition	Mean	SEM
Nonfamiliarized	1.27	0.17
4 Words	1.29	0.19
4 Sentences	1.19	0.18

S.6. Word Knowledge Assessment Performance for Children in Study 2

		Pretest		Immediate Posttest		Delayed Posttest	
Assessment	Condition	Mean	SEM	Mean	SEM	Mean	SEM
Spelling (Proportion Correct)	Control	0.30	0.05	0.39	0.05	0.49	0.06
	Nonfamiliarized	0.31	0.05	0.58	0.05	0.60	0.06
	4 Words	0.23	0.04	0.55	0.05	0.60	0.07
	4 Sentences	0.26	0.05	0.59	0.05	0.64	0.05
	Known Fillers	0.72	0.05	0.81	0.05	0.80	0.04
Perceived Familiarity (Max = 9)	Control	0.20	0.12	1.05	0.46	1.25	0.47
	Nonfamiliarized	0.75	0.19	1.34	0.29	2.15	0.51
	4 Words	0.81	0.24	2.53	0.40	2.00	0.49
	4 Sentences	0.39	0.15	1.45	0.37	1.93	0.49
	Known Fillers	7.93	0.29	7.71	0.29	8.00	0.31
Synonym Generation (Max = 3)	Control	0	0	0.19	0.11	0.04	0.04
	Nonfamiliarized	0.03	0.03	0.41	0.13	0.48	0.17
	4 Words	0	0	0.83	0.19	0.71	0.19
	4 Sentences	0.05	0.04	0.49	0.19	0.51	0.17
	Known Fillers	2.44	0.15	2.52	0.16	2.57	0.08
Synonym Matching (Proportion Correct)	Control	0.13	0.04	0.14	0.03	0.25	0.05
	Nonfamiliarized	0.19	0.04	0.49	0.06	0.39	0.05
	4 Words	0.18	0.05	0.50	0.07	0.40	0.07
	4 Sentences	0.21	0.04	0.50	0.07	0.46	0.07
	Known Fillers	0.84	0.04	0.87	0.03	0.83	0.04