CHILD DEVELOPMENT

Child Development, May/June 2020, Volume 91, Number 3, Pages 901–913

# The Production Effect Improves Memory in 7- to 10-Year-Old Children

Verena E. Pritchard, Michelle Heron-Delaney, and Stephanie A. Malone
*Australian Catholic University*

Colin M. MacLeod
*University of Waterloo*

The production effect—whereby reading words aloud improves memory for those words relative to reading them silently—was investigated in two experiments with 7- to 10-year-old children residing in Brisbane, Australia. Experiment 1 ($n$ = 41) involved familiar printed words, with words read aloud or silently appearing either in mixed- or blocked-list formats in a within-subject design. Recognition for words read aloud was better than for those read silently, an effect consistent across both list formats. These results were confirmed in Experiment 2 ($n$ = 40) using longer lists of printed novel nonwords. Final analyses indicated that the production effect was comparable for words and nonwords. Findings are discussed in relation to the distinctiveness account and the use of production as a mnemonic in children.

The production effect refers to the enhanced memory for items that are produced (vocalized) during encoding over those that are merely seen and read silently (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). This robust and replicable effect occurs for adult participants across a range of stimulus types, including words and nonwords, and under a variety of encoding conditions and retention intervals (Bodner, Jamieson, Cormack, McDonald, & Bernstein, 2016; Forrin, Jonker, & MacLeod, 2014; Icht, Mama, & Algom, 2014; Ozubko, Hourihan, & MacLeod, 2012; for a review, see MacLeod & Bodner, 2017). Better memory for produced words is also evident on a variety of tests, among them recognition, recall, and fill-in-the-blanks.

To date, the principal explanation of the production effect has been the distinctiveness account (e.g., MacLeod et al., 2010; Ozubko, Major, & MacLeod, 2014; cf. Jones & Pyc, 2014). Hunt (2006, p. 10) has described "*distinctive processing* as the processing of difference in the context of similarity"—essentially, noticing something unique during encoding that then is useful in aiding retrieval. Dodson and Schacter (2001; see also Israel & Schacter, 1997; Schacter, Israel, & Racine, 1999) have characterized this as a distinctiveness heuristic. Specifically, the act of production or vocalization (reading words aloud) during item-specific processing focuses attention on additional properties that provide for another element of encoding, an element that is useful later at the time of retrieval. Studies of the production effect have found that although other forms of production, such as mouthing, whispering, spelling, typing, and writing items, all improve memory relative to reading silently, vocalization at encoding remains the most effective type of production (Bodner et al., 2016; Forrin, MacLeod, & Ozubko, 2012; MacLeod et al., 2010; cf. Quinlan & Taylor, 2013).

To advance theoretical accounts, a substantial body of research has been dedicated to investigating the conditions that may attenuate or even eliminate the production effect. Foremost among these is the influence of experimental design. Although the effect is evident under both within-subject and between-subjects designs (Bodner et al., 2016; Fawcett, 2013; Forrin, Groot, & MacLeod, 2016), it is considerably more robust in within-subject designs, where aloud words and silent words typically are presented in a randomized mixed-list format, than in between-subjects designs, where each of the two conditions is presented to a separate group of participants (e.g., Hopkins & Edwards, 1972; Jonker, Levene, & MacLeod, 2014; MacLeod et al., 2010).

This research was supported by the Australian Catholic University Under-Graduate Research Internship Program. The authors specially thank Dr Kate Witteveen for assistance with stimuli preparation and to Tiffany World and Clare Noonan for data collection. The authors also thank the schools and children involved for their support of this project. The authors have no potential conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Verena E. Pritchard, Institute of Learning Sciences and Teacher Education, Australian Catholic University, Level 4, 229 Elizabeth St, Brisbane, Queensland 4000, Australia. Electronic mail may be sent to verena.pritchard@acu.edu.au.


DOI: 10.1111/cdev.13247

This pattern fits with the distinctiveness account, in that distinctiveness ordinarily will be much more salient in a mixed-list situation. As a mnemonic, the memorial benefits of vocalization are widespread, occurring even in older adults (Lin & MacLeod, 2012), an age group typically characterized by costs in memory monitoring and deficits in retrieving distinctive information (Carr, Castel, & Knowlton, 2015).

The ability to recall or to recognize words accurately has clear benefits for learning, particularly for younger age groups where language and communication skills are critical to development across a range of domains including motor (Horn, Pisoni, & Miyamoto, 2006) and behavioral (Petersen et al., 2013) regulation, as well as learning (Snowling & Hulme, 2006). Despite this, the production effect remains surprisingly unexplored in children, particularly for printed items—the situation most extensively studied in adults. Consequently, the primary goal of this study was to investigate the production effect for familiar and novel verbal materials in young children of reading age, following the general procedural design of MacLeod et al. (2010; experiments 3A and 6).

To date, only two studies have examined production effects in children (Icht & Mama, 2015; Zamuner, Strahm, Morin-Lessard, & Page, 2017), both in preschool children and hence in nonreaders. Thus, both studies used a combination of oral and pictorial stimuli that did not require reading. In the earlier study, Icht and Mama used a within-subject design (thirty 5-year-old children per experiment) to assess recall for pictures of familiar objects (Experiment 1) and recognition for unfamiliar objects (Experiment 2). Familiar objects were defined as those representing high-frequency nouns likely to exist in the expressive lexicon of 5-year-olds (e.g., hammer, tiger). Unfamiliar objects were those representing low-frequency (or rare) nouns (e.g., pestle, manager). There were three conditions: (a) produced ("look and say"), (b) heard ("look and listen"), and (c) seen ("look"). In both experiments, participants viewed a total of 30 objects, 10 per condition. During the study phase, flashcards depicting each object were randomly placed into one of three boxes, with each box indicating the appropriate condition-dependent response (i.e., produce, listen, or look) to the child. Words corresponding to each condition were randomly intermixed (i.e., a mixed-list design). In the same session, children were asked either to recall aloud as many objects as they could remember (Experiment 1) or to identify the studied objects on a four-alternative forced choice recognition test (Experiment 2). Icht and Mama found enhanced memory for items in the production condition in both experiments, indicating a reliable production effect in preschool children. They attributed this effect to the articulatory, phonetic, and sensory-motor processes invoked during encoding.

Zamuner et al. (2017) took a different tack, examining whether children would show a recognition advantage for produced novel nonwords that had no preexisting semantic representations. In their study of 16 children aged 4.5–6 years, spoken nonwords were used as stimuli, with a novel animal-like drawing assigned to each. In their first experiment, children were presented with two sets of four novel phonotactically legal consonant-vowel-consonant nonwords, one set assigned to a produced condition and one to a heard condition. Prior to the study phase, the children were trained to respond to a prompt image appearing below the animal referent that indicated whether they should listen to the nonword in silence (heard condition) or say it aloud (production condition) when they saw its assigned animal referent. During the study phase, children heard the nonword and saw its corresponding visual referent. Each nonword was presented on two study trials with the nonwords associated with each condition presented using a within-subject, mixed-list design. In the recognition test phase, the children saw all four visual referents in turn, each appearing with a distractor item. Surprisingly, on a preferential looking test, Zamuner et al. reported that a "reverse production effect" was observed, with children's looking times for the nonword targets longer in the heard condition than in the produced condition.

To investigate whether their finding of enhanced memory for heard rather than produced items related to potential methodological and design issues, Zamuner et al. (2017) ran a second experiment increasing the number of nonwords in the study phase from four to eight. These were presented in a within-subject blocked-list format rather than in a mixed-list format so that children studied a block of silent nonwords followed by a block of aloud nonwords or vice versa. In addition to measuring preferential looking, a free recall task was also included where participants simply reported the names of any of the animals that they could remember. Despite these changes, children still showed better recognition (and recall) for nonwords that had been heard rather than produced. Zamuner et al. concluded that the production effect in children may reverse or attenuate when novel stimuli with nonnative phonemes or with infrequent sound patterns are used and also when the task is cognitively challenging. Specifically, a task that requires children to first map the auditory renditions of novel items to visual referents was argued to have absorbed the cognitive resources that would otherwise have been available for learning, the result being a memorial cost for items in the produced condition.

Thus far, these are the only two studies of the production effect in children. Both have studied preschool children and consequently have used pictorial stimuli that have required the cognitively demanding task of sound-to-object mapping. Given that the usual mode of vocabulary acquisition during formal schooling uses the printed form, there is a clear need to establish whether there is a benefit of production for printed rather than pictorial content in children. This will, of course, require testing of school-age children who can read. Furthermore, there have been no systematic attempts to map out conditions that influence the degree to which the effect occurs in children.

As alluded to earlier, there is substantial evidence (see MacLeod & Bodner, 2017) to indicate that production effects are considerably larger in within-subject mixed-list designs (in which a random 50% of words are read aloud and 50% are read silently) than in between-subjects blocked-list designs (in which one group produces 100% of words and the other produces 0%). This finding is consistent with the relative distinctiveness explanation (however, see Fawcett, 2013, for a review; Jones & Pyc, 2014), in that in the between-subjects situation there is no distinction during study between mode of processing items. The much smaller production effect observed when aloud–silent is manipulated between-subjects may arise from one or more of three possible sources: (a) strengthening by production, a small increment that occurs regardless of design (see MacLeod & Bodner, 2017), (b) absence of the cost to silent items that occurs in within-subject mixed lists (Bodner et al., 2016), and/or (c) use of a distinctiveness strategy at test ("Did I say this aloud?"; see Dodson & Schacter, 2001) by only a minority of participants in the aloud group or on only a subset of test items because the potential value of this strategy at the time of test is much less apparent when all items were studied in the same way.

If there are design-contingent patterns in younger school-age children, this could have practical implications regarding the use of production as a mnemonic in educational or clinical settings. For example, is it best for target words to be mixed or kept together in terms of whether they are read aloud or silently during learning? Even with adult participants, few studies have investigated within-subject effects for blocked lists versus mixed lists (Bodner, Taikh, & Fawcett, 2014; Bodner et al., 2016; Icht et al., 2014); the majority are restricted to comparisons of a within-subject mixed-list design to a between-subjects pure-list design (see Forrin et al., 2016; Hopkins & Edwards, 1972; MacLeod et al., 2010) and the influences of statistical distinctiveness (i.e., comparing effects for the ratios of 20:80, 50:50, and 80:20 aloud vs. silent). Comparing within-subject production effects in children across blocked-list and mixed-list formats will clarify whether the production advantage in this age group is influenced by list type and hence how this mnemonic should best be implemented to maximize learning and remembering.

Separating aloud and silent items by blocking at study may, for example, reduce the distinctiveness of aloud items at encoding or participants' use of the "distinctiveness heuristic" at test. However, at least in adults, the evidence suggests this may not be the case. In one of the few studies to have examined this, Bodner et al. (2014) compared adults' recognition rates across aloud and silent items for within-mixed versus within-blocked designs and reported equivalent recognition rates for aloud items across the two designs. The same was true of silent items, eliminating lazy or cursory reading of unproduced items as a potential source of production effects.

This was not true, however, of the comparisons between within-mixed versus pure silent conditions implemented by Bodner et al. (2014), where silent items suffered a relative within-subject cost. These cost/benefit analyses indicate that the distinctiveness-based and the strength-based accounts of the production effect are not necessarily mutually exclusive: Both may be operating (see also Bodner et al., 2016; MacLeod & Bodner, 2017; Fawcett & Ozubko, 2016). Under a strength account, production results in a benefit both in pure lists and in mixed lists. In mixed-lists, however, production enhances the overall strength of a produced item's representation via enhanced familiarity but this gain does not occur for silent items (see Fawcett & Ozubko, 2016; Ozubko, Gopie, & MacLeod, 2012). The potential for differences in recognition rates for aloud words across within-subject mixed-list versus blocked-list designs has not been examined in children. Delineating any such design-contingent patterns of the production effect in children is potentially important for the purpose of establishing the practical use of vocalization as a mnemonic to enhance word learning at early school age. To this end, a secondary focus of this study was to compare the within-subject production effects for blocked lists and mixed lists in children.

Using the general procedure of MacLeod et al. (2010), where printed words are used as stimuli, the production advantage has been repeatedly demonstrated in young adults. Our principal aim here was to evaluate the production effect using these more typical materials (i.e., printed words) and following the same procedure. Thus, our study included children aged 7- to 10-years-old, an age group where growth in reading fluency undergoes its most rapid development and for whom reading is becoming well established (Hill, Bloom, Black, & Lipsey, 2007; Logan et al., 2013). Further aims were: (a) to compare effects for familiar and unfamiliar verbal stimuli to reexamine the Zamuner et al. (2017) hypothesis that when stimuli are infrequent or nonnative, an attenuated or reversed production effect is observed; and (b) to determine the effect of blocking on the within-subject production effect by comparing the effect across list type (mixed vs. blocked).

## Experiment 1: Familiar Words

In this first experiment, closely following the MacLeod et al. (2010) procedure, we examined the production effect for familiar printed words using a recognition test. We studied children early in their school years, specifically children aged 7- to 10-years-old, because they have just recently begun to read. Consequently, this experiment is the first to explore whether the production advantage occurs in school-age children, as well as the first to use printed verbal materials with children. To determine any potential influence on the production effect of how the two conditions (aloud vs. silent) were presented, we directly compared mixed-list versus blocked-list procedures, the former randomly intermingling words in the two conditions and the latter presenting all of one condition before all of the other condition. Examining the potential influence of mixing versus blocking on the production effect also allows us to compare the effectiveness of two different learning strategies employed in the rote memorization of list words.

### *Method*

#### *Participants*

Forty-one children (21 female, 20 male) were recruited from 3 to 5 years at an independent primary school in Brisbane, Australia. An opt-in procedure was followed. According to Government data on the socioeconomic composition of this school population (Index of Community Socio-Educational Advantage; ICSEA), this school has an ICSEA rating of 952. This score reflects a student population with a slightly below average level of educational advantage (ICSEA average is 1,000; $SD = 100$). The ethnic composition of the school includes indigenous students (11%) and students with a language background other than English (15%). Data were collected during August and September in 2017. The mean age of participants was 9.07 years old ($SD = 0.95$; range = 7.10–10.11). Children were randomly assigned to one of two groups. The Mixed List group ($n = 20$) included 9 females and 11 males with a mean age of 9.06 ($SD = 0.87$; range = 8.03–10.11). The Blocked List group ($n = 21$) included 9 males and 12 females with a mean age of 9.08 ($SD = 1.05$; range = 7.10–10.11). All participants were identified by their teachers as being fluent in English and as having reading ability satisfactory for completing the study. Children were tested following written informed parental consent; the children also provided verbal assent prior to testing. The study protocol was approved by the Human Research Ethics Committee at Australian Catholic University (Ref: 2016-212H).

#### *Materials*

The stimuli, shown in Appendix A, consisted of 120 words (e.g., *school, forest, dinner*) sourced from MacLeod et al. (2010). These were all nouns, 5 to 10 letters in length and similar to the words corresponding to the pictorial stimuli used by Icht and Mama (2015). Word difficulty was deemed age-appropriate based on comparison with the standardized Single Word Reading Test for children (York Assessment of Reading; Snowling et al., 2009). From this pool, 80 words were randomly selected for the study and recognition phases of the experiment. These words were then randomized again and color coded (40 in blue and 40 in orange) to indicate whether they were to be spoken aloud or read silently. Twenty words of each color were then randomly assigned to the study phase. The colors assigned to the two conditions (aloud vs. silent) were counterbalanced: For 50% of participants, blue denoted read aloud; for the remaining 50%, orange denoted read aloud. For participants in the Mixed List group, words designated as "aloud" or "silent" were randomly intermixed. For participants in the Blocked List group, the 20 aloud words appeared consecutively, as did the 20 silent words, with the presentation order of the aloud versus silent blocks also counterbalanced. All words were presented via PowerPoint on a laptop, with the colored text (44-character lowercase) centered on a white background.

### *Design and Procedure*

The procedure was modeled on that of MacLeod et al. (2010). A mixed design was used. The within-subject variable was Item Type (aloud vs. silent); the between-subjects variable was List Type (mixed vs. blocked). Participants in both the Mixed-List and Blocked-List groups were tested individually in a quiet room, either at school or in their home, and seated in front of a laptop that displayed the stimuli. For the study phase, words were presented individually for 2,000 ms with each word separated from the next by a white screen displayed for 1,000 ms. The children were instructed to read the words appearing in a particular color (e.g., blue) aloud and those in the other color (e.g., orange) silently, depending on the color preassigned to the condition. Each presentation in the study phase began with 1,000 ms of white screen that preceded the appearance of the first word. Immediately following the study phase and the test instructions, a recognition test was administered. Participants were presented with the 40 studied words together with 20 unstudied words shown individually in a new random order in black font at the center of a white screen. Words in both the study and test phases were presented in a new random order for each child. Participants were asked to respond aloud "yes" if they remembered the word from the study phase and "no" if they did not. The experimenter manually recorded all responses.

## *Results*

### *False Alarms*

An independent samples *t*-test revealed no significant difference in false alarm rates—saying "yes" to new distractor words—between children in the Mixed-List group ($M$ = .59, $SE$ = .07) and the Blocked-List group ($M$ = .64, $SE$ = .08), $t(39)$ = .48, $p$ = .63. Consequently, subsequent analyses were restricted to hit rates rather than using a signal detection measure of memory discrimination ($d'$), given the demonstrated disadvantages of using $d'$ with a within-subject design (see Forrin et al., 2016) where only a single false alarm rate is available.

### *Overall Hits Versus False Alarms*

Mean hit rates (yes responses to studied items; aloud plus silent) for the Mixed-List group ($M$ = .78, $SE$ = .03) and the Blocked-List group ($M$ = .78, $SE$ = .04) did not differ; $t(39)$ = .82, $p$ = .42. Overall, mean proportion of false alarms was well below that of hits (.62 vs. 76), demonstrating that the children were able to distinguish correctly between studied (old) and unstudied (new) familiar words. The false alarm rate was higher than is typically seen in adults (.62 here vs. .19 in experiment 1A of MacLeod et al. 2010), suggesting a possible "yes" bias in children.

### *The Production Effect*

The upper section of Table 1 presents the hit rates for words studied aloud versus silently. The hit rates were analyzed in a 2 (list type: mixed vs. silent) × 2 (item type: aloud vs. silent) mixed analysis of variance (ANOVA). There was a significant main effect of Item Type indicating a reliable production effect, with hit rates for aloud items ($M$ = .84, $SE$ = .04) consistently higher than for silent items ($M$ = .68, $SE$ = .04), $F(1, 39)$ = 25.96, $p$ < .001. There was no influence of List Type, with hit rates for blocked items ($M$ = .81, $SE$ = .04) similar to those for mixed items ($M$ = .87, $SE$ = .03), $F$ < 1. And finally, there was no evidence that the observed production effect differed between the mixed-list and blocked-list groups, $F$ < 1. The results of a 2 (list type: mixed vs. blocked) × 2 (item type: aloud vs. silent) × 2 (gender: male vs. female) ANOVA showed no significant main effect or interaction involving gender either before ($p$ = .21) or after ($p$ = .13) covariate adjustment for age.

### *Block Order*

A further analysis explored the possibility of a block order interaction within the Blocked-List group using a 2 (block order: silent–aloud vs. aloud–silent) × 2 (item type: aloud vs. silent) mixed ANOVA. There was a significant effect of Item Type—the production effect, $F(1, 20)$ = 6.61, $p$ = .02.

Table 1
*Mean Proportions of Hits (Correct Yes Responses) and Standard Errors as a Function of Study Condition and List Design*

| | Aloud | | Silent | | Collapsed | |
|---|---|---|---|---|---|---|
| | *M* | *SE* | *M* | *SE* | *M* | *SE* |
| Experiment 1 | | | | | | |
| Mixed-list group | .87 | .03 | .70 | .04 | .79 | .04 |
| Blocked-list group | .81 | .04 | .67 | .05 | .74 | .04 |
| Collapsed | .84 | .04 | .68 | .05 | .76 | .04 |
| Silent–aloud group | .80 | .06 | .63 | .05 | | |
| Aloud–silent group | .82 | .03 | .71 | .04 | | |
| Experiment 2 | | | | | | |
| Mixed-list group | .70 | .03 | .53 | .03 | .62 | .03 |
| Blocked-list group | .67 | .04 | .54 | .04 | .60 | .04 |
| Collapsed | .68 | .04 | .54 | .04 | .61 | .04 |
| Silent–aloud group | .71 | .04 | .54 | .04 | | |
| Aloud–silent group | .64 | .05 | .55 | .03 | | |

There was, however, no main effect of Block Order, $F < 1$, nor was there a Block Order by Item Type interaction, $F < 1$, indicating that the production effect was comparable across the two block orders. The means and standard errors for the silent–aloud order participants ($n = 10$) and the aloud–silent order participants ($n = 11$) are displayed in Table 1.

### *Discussion*

Experiment 1 clearly showed a reliable production effect for familiar printed words in school-aged children. This was evident and comparable regardless of whether their study condition used a mixed-list or a blocked-list procedure. This replicates the pattern found for adults by MacLeod et al. (2010) using a within-subject mixed-list design. As such, this is the first experiment to establish a typical production effect for printed items in children of reading age. Also of note, the hit rates observed for aloud items (.87) and for silent items (.70) in the mixed-list group are similar to those observed for young adults (.87 and .74, MacLeod et al., experiment 3A). Corroborating the results of Icht and Mama (2015) for preschool children with pictorial stimuli, production appears to be a useful encoding tactic for younger age groups. Collectively, these results present further evidence to support the idea that the production advantage is a robust and highly replicable phenomenon seen in children, young adults, and older adults. Experiment 1 also showed no effect of mixing versus blocking on the within-subject production effect, indicating that production enhances memory independently of the learning strategy used by the child or educator.

## Experiment 2: Novel Nonwords

Experiment 1 showed a robust memory advantage for familiar words read aloud over those read silently in 7- to 10-year-old children. This effect held across both mixed-list and blocked-list conditions, assessed in two independent samples. In Experiment 2, we moved on to examine recognition rates for novel (unfamiliar) printed nonwords for similarly aged children. MacLeod et al. (2010) had reported a large production effect for nonwords. The primary purposes of Experiment 2 were to replicate and generalize the findings of Experiment 1 and to evaluate the Zamuner et al. (2017) claim that the production effect is reversed in children when infrequent or nonnative stimuli are used. An additional aim was again to test the replicability of our finding of no effect of mixing versus blocking on the within-subject production effect.

In Experiment 2, we also doubled the number of items in the study phase, a manipulation that resulted in list lengths more typical of the studies using adults as participants (see MacLeod et al., 2010), thereby providing further generalizability.

### *Method*

#### *Participants*

A new sample of 40 children was recruited from the same source as in the previous experiment. The mean age of participants was 8.92 years old ($SD = 0.80$; range = 7.20–10.70). As before, children were randomly allocated to either the Mixed-List group or the Blocked-List group. Participants in the Mixed-List group ($n = 20$) included 11 males and 9 females. The mean age of participants in this group was 8.91 years old ($SD = 0.92$; range = 7.20–10.70). The Blocked-List group ($n = 20$) included 8 males and 12 females, with a mean age of 9.08 years ($SD = 1.05$; range = 7.10–10.11).

#### *Stimuli*

The stimuli, shown in Appendix B, were a set of 120 pronounceable nonwords (e.g., *hest*, *slass*, *prench*) sourced from MacLeod et al. (2010). All were monosyllabic, four to six letters in length, and ended in consonants. These nonwords were readily pronounceable on inspection by teachers, and therefore deemed appropriate for 7- to 10-year-old children. From this pool, 80 nonwords were randomly selected for the study phase of the experiment. These were then randomized again and color coded so that 40 appeared in blue and 40 in orange during the study phase. This larger stimulus set increased cognitive demand during the study phase and matched the list length used by MacLeod et al. (2010). The colors assigned to read aloud versus silent words were counterbalanced. For participants in the Mixed-List group, the presentation order of aloud versus silent stimuli was randomized across participants. For participants in the Blocked-List group, the 40 aloud words appeared consecutively, as did the 40 silent words, with block order counterbalanced. The recognition test phase included 40 studied (20 per condition chosen at random) and 20 unstudied stimuli presented in black, with selection randomized across participants. Selecting only 40 (of the 80) studied words made the length of the recognition test phase comparable to that of Experiment 1.

### *Design and Procedure*

As before, a 2 (list type: mixed vs. blocked) × 2 (item type: aloud vs. silent) mixed design was used, emulating Experiment 1. All procedural aspects were identical to those in Experiment 1 with the exception that at study children were presented with 40 nonwords per condition instead of the 20 familiar words per condition used in Experiment 1.

## *Results*

The lower section of Table 1 shows the yes/no recognition data expressed as proportions of *yes* responses for Experiment 2.

### *False Alarms*

Again, an independent samples *t*-test was first conducted to compare false alarm rates between the two study groups (Mixed-List vs. Blocked-List). There was no significant difference in false alarms between children in the Mixed-List ($M = .29$, $SE = .04$) and the Blocked-List groups ($M = .27$, $SE = .04$) groups, $t(38) = .24$, $p = .81$. Consequently, subsequent analyses on the production effect were restricted to hit rates. It is worth noting that false alarm rates were much lower in Experiment 2 than in Experiment 1, suggesting that the yes bias for words did not extend to nonwords.

### *Overall Hits Versus False Alarms*

Mean hits (silent plus aloud) for the Mixed-List ($M = .62$, $SE = .03$) and Blocked-List ($M = .61$, $SE = .04$) groups were also compared. No significant difference was evident between the groups, $t$ $(38) = .24$, $p = .81$. Again, the mean proportion of false alarms was well below that of hits (.28. vs. .61), demonstrating that children were able to distinguish correctly between studied and unstudied (new) nonwords.

### *The Production Effect*

Hit rates were subjected to a 2 (list type: mixed vs. silent) × 2 (item type: aloud vs. silent) mixed ANOVA. There was a significant main effect of Item Type indicating a typical production effect, with hit rates for aloud items ($M = .69$, $SE = .04$) higher than for silent items ($M = .54$, $SD = 0.04$), $F$ $(1, 38) = 64.76$, $p < .001$. There was no main effect of List Type, $F < 1$, with performance of the Blocked group ($M = .67$, $SE = .04$) very similar to that of the Mixed group ($M = .70$, $SE = .02$). No interaction was evident, $F(1, 38) = 1.32$, $p = .23$, indicating that the production effect for nonwords was comparable across the Mixed-List and Blocked-List groups. The results of a 2 (list type: mixed vs. blocked) × 2 (item type: aloud vs. silent) × 2 (gender: male vs. female) ANOVA showed no significant interaction either before ($p = .41$) or after ($p = .45$) covariate adjustment for age.

### *Block Order*

To check for any possible influence of block order on the production effect, we carried out a 2 (List order: silent–aloud vs. aloud–silent) × 2 (item type: aloud vs. silent) mixed ANOVA. There was a significant production effect, $F(1, 19) = 27.06$, $p < .001$, with recognition better for aloud ($M = .67$, $SE = .04$) than for silent ($M = .54$, $SE = .04$) nonwords. There was no main effect of block order, $F < 1$, and no reliable item x order interaction, $F(1, 19) = 3.01$, $p = .10$, indicating that the production effect was similar across the two block orders. The means and standard errors for the silent–aloud order participants ($n = 10$) and the aloud–silent order participants ($n = 10$) are shown in Table 1.

### *Cross-Experiment Comparison*

To examine consistency across item types, we carried out one further analysis on hit rates: a combined 2 (item type: aloud vs. silent) × 2 (list type: mixed vs. blocked) × 2 (experiment: 1 vs. 2) mixed ANOVA, where experiment was the additional between-subjects variable. There was of course a significant main effect of Item Type, $F(1, 80) = 72.19$, $p < .001$, with aloud ($M = .76$, $SE = .02$) greater than silent ($M = .61$, $SE = .02$). As well, there was a significant effect of Experiment, $F(1, 80) = 19.77$, $p < .001$, with recognition performance lower in Experiment 2 ($M = .69$, $SE = .03$) than in Experiment 1 ($M = .84$, $SE = .03$), no doubt because the nonwords were unfamiliar and hence harder to remember, possibly also reducing any yes bias demonstrated for words in Experiment 1. The main effect of List Type was not significant, $F < 1$. Importantly, none of the interactions were significant, $F$s < 1, $p$s > .33, indicating that the production effect was entirely consistent across both list types and stimulus types.

### *Discussion*

Experiment 2 demonstrates for the first time that the production effect is observed in children even when novel nonword stimuli are used. That the effect is not limited to familiar words, or indeed to words at all, suggests that it could be of benefit in learning new, unfamiliar words as well. These results replicate those obtained with familiar words in Experiment 1 and highlight the generalizability of the effect across different stimulus types in children, just as has been observed in adults (MacLeod et al., 2010).

The results of Experiment 2 differ from the Zamuner et al. (2017) report of a reverse production effect—better memory for silent than for aloud items—and hence also their conjecture that production effects in children may be dependent on stimulus type and be unlikely to occur for stimuli that are nonnative or infrequent. (Of course, this difference in pattern may hinge on the difference in materials or in age group between our study and theirs, as well as on the fact that Zamuner et al.'s study also included a *heard* condition). Rather, the production advantage in children appears to be very robust. Even with double the number of stimuli at study, the hit rates for children here (.70 aloud and .53 silent) were highly consistent with those observed for adult participants by MacLeod et al. when the materials were nonwords (2010; experiment 6; .71 aloud and .52 silent). Similar to effects for familiar words, neither list type nor block order (i.e., aloud–silent first vs. silent–aloud first) influenced the production effect for nonword items. This is the first report of a within-subject mixed-list versus blocked-list examination of the production effect in children; indeed, relatively little work has explored this issue in the adult literature.

Collectively, the results of Experiment 2 are entirely consistent with those of Experiment 1, suggesting that the production advantage is a highly replicable and robust effect in children, even for novel nonwords. Our final set of analyses further showed that the production effect is not influenced by list type for this age group, and that it is comparable for words and nonwords. These conclusions also hold up regardless of study list length. In short, production appears to be an effective mnemonic for pronounceable novel nonwords that have neither meaning nor prior memory representation, just as it is for common real words. Thus, production has the potential to be valuable in word learning, in that nonwords can be viewed as words that are not yet familiar, and may be more easily learned and remembered when spoken aloud.

## General Discussion

The experiments reported here are the first to have examined the production effect in school-age children. Here, we investigated the production effect in 7- to 10-year-old children using both familiar words and novel nonwords. Production effects and within-condition recognition rates were compared across both mixed-list and blocked-list within-subject designs using independent samples both to determine the generalizability of the production effect and to test for any influence of list type (mixed-list vs. blocked-list formats). Across two experiments, we have shown that the production advantage is a robust phenomenon in young school-age children—indeed, apparently as robust as it is for adults—and that it is not limited to stimuli with existing lexical or semantic representations nor is it restricted to a particular list presentation format.

### *Production As a Mnemonic for Children*

Following the procedural design of MacLeod et al. (2010: experiments 3A and 6), Experiments 1 and 2 here successfully extended effects found with adults to school-age children, demonstrating age continuity of the benefit of oral production. This was true both for familiar words and for unfamiliar nonwords. Moreover, we observed rates of recognition for both aloud and silent words in these children that were remarkably consistent with those previously observed for adults. These results accord well with literature suggesting that the production effect is a robust and replicable phenomenon occurring for older adults (Lin & MacLeod, 2012) as well as for younger adults (MacLeod et al., 2010; Quinlan & Taylor, 2013; see MacLeod & Bodner, 2017, for a review). Unlike findings with older adults, where some attenuation of the effect has been observed relative to younger adults, consistent with their reduced distinctiveness of encoding, our effects for children were very consistent with those in young adults for both familiar and novel stimuli. This suggests that speech production is an effective encoding tactic for school-age children and that this age group would appear to have less difficulty in retrieving distinctive information than is the case for older adults (see Geraci, McDaniel, Manzano, & Roediger, 2009). This pattern is also consistent with evidence showing intact source memory for children by early school age (Hayne & Imuta, 2011; Newcombe, Lloyd, & Ratliff, 2007).

In contrast to Zamuner et al. (2017), we found no evidence for a reverse production effect in children —or even for an attenuated effect—when novel items rather than familiar items comprised the studied material. Given our simple task where cognitive demands are relatively low, it seems that production can be a highly effective mnemonic for children of early- to mid-school age during word learning. This is not to say that the effectiveness of this mnemonic for encoding unfamiliar stimuli is not dependent on factors such as task demands or complexity, particularly at younger ages. As suggested by Zamuner and colleagues, better recognition rates for unfamiliar (or nonnative) verbal stimuli that are heard rather than produced at encoding could indicate that passive listening may be a better learning strategy for very young children. Specifically, for these young children, listening to, rather than producing, novel speech sounds may free up the necessary cognitive resources for the creation of new phonological, lexical, and semantic representations. This account accords well with research indicating that phonological short-term memory capacity constrains word learning in 4-year-old children (Gathercole, Service, Hitch, Adams, & Martin, 1999).

Additionally, superior phonological memory function is also associated both with more rapid learning of the phonological aspects of novel words (Gathercole & Baddeley, 1990; Gathercole et al., 1999; Michas & Henry, 1994; Papagno & Vallar, 1992) and with an increased ability to acquire foreign vocabulary (Papagno & Vallar, 1995; Service, 1992; Service & Kohonen, 1995). Tasks such as that of Zamuner et al. (2017) that require children to map sets of auditorily presented novel verbal stimuli to visual referents prior to production are likely to load heavily on phonological working memory, thereby possibly interfering with speech output as well as learning, perhaps especially when that learning rests on production. Lexical retrieval and word finding difficulties have been documented for children when pronunciation is affected (Messer & Dockrell, 2006).

In using a yes/no recognition test, rather than a free recall test, we may have further reduced some of the potential cognitive demands. Production effects might not have been as evident for the novel nonwords had a free recall task been used, given the associated increase in cognitive demand and the strength of the phonological representation required (Litt & Nation, 2014). This represents an avenue for further research. Exploring longer retention intervals and the durability of effects will also be useful to probe the relevance of the production paradigm for lexical acquisition. As a final point here, relative to older children, it may also be that preschool children require a more concrete situation for production to aid memory, with their word learning more dependent on contextualized mappings between the word and the world (Dockrell & Messer, 2004). For example, the production effect may only be evident in younger preschool children when presented with words that they have already encountered to some degree, therefore enabling them to use their already existing lexicon-visual object mapping. The encoding of novel words may also only be partial for younger children resulting in a weak phonemic representation (e.g., the non-word *hest* encoded as *he*). As demonstrated by Icht and Mama (2015), engaging in production certainly does appear to assist preschool children in learning familiar pictorial stimuli (i.e., their "look and say" condition).

The simple nature of the production technique lends itself well for use in educational settings, as Ozubko, Gopie, et al. (2012) and Ozubko, Hourihan, et al. (2012) have demonstrated with text materials. The fact that school-age children's learning was incidental in our study (i.e., we did not inform children that a memory test would follow the study phase) indicates that speech production results in enhanced learning with a reasonable level of durability even without intention to learn. Given that vocabulary knowledge is a well-established predictor of academic success (Cunningham & Stanovich, 1997; see also Moghadam, 2012, for a review), mnemonic aids suitable for young children are important in scaffolding literacy development. Survey-based research has indicated that teachers perceive 10%–16% of children to have learning difficulties in literacy at early school age (Butterworth & Kovas, 2013; Louden et al., 2006). Matthew effects in education (Stanovich, 2017), whereby initially modular learning deficits broaden into a global difficulty, also are not uncommon (Ceci, 1991; Share & Silva, 1987; Stanovich, 1993). To minimize the impact that early learning challenges can have on educational attainment and learning motivation, the use of easily implemented mnemonic techniques in early word learning should be taught and encouraged.

### *Developmental and Further Considerations*

Finally, this is the first study to examine the potential influence of different presentation conditions—specifically mixing versus blocking conditions—on the within-subject production effect in children, in the hope of establishing which item presentation formats may be most likely to be beneficial for children to learn and remember. Comparing the within-subject production effect across mixed lists and blocked lists, we found that presentation format did not matter for either familiar words or unfamiliar nonwords in children. Equivalent effects across the two list formats provide evidence to suggest that production works well for learning in children regardless of whether they (or a relevant educator) choose to separate or intermix the material to be learned. This is consistent with the idea that production accentuates distinctiveness when other items are not produced. Children in the mixed-list and blocked-list groups appeared equally likely to use distinctiveness, whether during encoding, as a heuristic at test, or both.

Clearly there is a need for further research on this issue, including studies with larger sample sizes. Nevertheless, our experiments indicate that production of some items did not appear to impair children's memory for unproduced items in the same list. Considering the use of this mnemonic in clinical or educational settings, grouping target items may be equally effective to mixing them: It is the aloud-silent contrast—the distinctiveness—that is crucial. Also worthy of further investigation is the apparently greater false alarm rates in children than in adults, particularly when familiar words were used in Experiment 1. False alarm rates were lower in Experiment 2 where nonwords were used, suggesting greater recognition accuracy for these items. This difference may be due to the increased novelty of nonwords relative to familiar words, where a bias to say "yes" may exist based on preexperimental familiarity of the words.

Relatedly, future research should determine whether production may be even more effective in enhancing recognition for nonwords relative to familiar words due to increased use of the distinctiveness heuristic (see Dodson & Schacter, 2001) for such items during learning, particularly among younger than older age groups. Future studies should also consider the extent to which the production effect may vary according to the skill level of the reader. Doing so may help to highlight whether reading skill (in addition to age) should be considered as an important covariate in future studies of the production effect for words and nonwords in studies involving children. Finally, although the production effect was comparable across the two experiments, it should be acknowledged that differences in item type (words vs. nonwords) and study list length (shorter vs. longer) may have influenced the results that we observed. Hence, future studies evaluating the consistency or reliability of the production effect for blocked- versus mixed-list designs in children should keep both stimulus type and list length constant.

To conclude, production does appear to significantly assist 7- to 10-year-old children of early school age in remembering. The production effect that we observed in these children was robust across both familiar and unfamiliar printed material and was not restricted to a mixed-list design. Focus should now shift to establishing the influence of different retrieval conditions under which the effect is evident in this age group, as well as to investigating the influence of production on other educationally relevant materials, such as texts (see Ozubko, Gopie, et al., 2012; Ozubko, Hourihan, et al., 2012). Further work examining whether production enhances memory for associative information in addition to item information (see Putnam, Ozubko, MacLeod, & Roediger, 2014) will also be useful given the relevance of associative learning in the acquisition of language and reading skills (Mourgues et al., 2016). Such a program of research will shed light on the different aspects of learning and memory that are enhanced by speech production in children.

## References

Bodner, G. E., Jamieson, R. K., Cormack, D. T., McDonald, D. L., & Bernstein, D. M. (2016). The production effect in recognition memory: Weakening strength can strengthen distinctiveness. *Canadian Journal of Experimental Psychology*, *70*, 93–98. https://doi.org/10.1037/cep0000082

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition.

*Psychonomic Bulletin & Review*, *21*, 149–154. https://doi.org/10.3758/s13423-013-0485-1

Butterworth, B., & Kovas, Y. (2013). Understanding neurocognitive developmental disorders can improve education for all. *Science*, *340*, 300–305. https://doi.org/10.1126/science.1231022

Carr, V. A., Castel, A. D., & Knowlton, B. J. (2015). Age-related differences in memory after attending to distinctiveness or similarity during learning. *Aging, Neuropsychology, and Cognition*, *22*, 155–169. https://doi.org/10.1080/13825585.2014.898735

Ceci, S. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*, 703–722. https://doi.org/10.1037/0012-1649.27.5.703

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*, 934–945. https://doi.org/10.1037/0012-1649.33.6.934

Dockrell, J. E., & Messer, D. (2004). Lexical acquisition in the early school years. In R. A. Berman (Ed.), *Language development across childhood and adolescence* (pp. 35–52). Amsterdam, The Netherlands: John Benjamins. https://doi.org/10.1075/tilar

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155–161. https://doi.org/10.3758/BF03196152

Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, *142*, 1–5. https://doi.org/10.1016/j.actpsy.2012.10.001

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-lists production effect in recognition memory. *Canadian Journal of Experimental Psychology*, *70*, 99–115. https://doi.org/10.1037/cep0000089

Forrin, N. D., Groot, B., & MacLeod, C. M. (2016). The d-prime directive: Assessing costs and benefits in recognition by dissociating mixed-list false alarm rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1090–1111. https://doi.org/10.1037/xlm0000214

Forrin, N. D., Jonker, T. R., & MacLeod, C. M. (2014). Production improves memory equivalently following elaborative vs. non-elaborative processing. *Memory*, *22*, 509–524. https://doi.org/10.1080/09658211.2013.798417

Forrin, N. D., MacLeod, C. M., & Ozubko, J. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*, 1046–1055. https://doi.org/10.3758/s13421-012-0210-8

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*, 336–360. https://doi.org/10.1002/(SICI)1099-0720(199902)13:1<65:AID-ACP548>3.0.CO;2-O

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, *13*, 65–77. https://doi.org/10.1002/(sici)1099-0720(199902)13:1%3c65::aid-acp548%3e3.0.co;2-o

Geraci, L., McDaniel, M. A., Manzano, I., & Roediger, III, H. L. (2009). The influence of age on memory for distinctive events. *Memory & Cognition*, *37*, 175–180. https://doi.org/10.3758/MC.37.2.175

Hayne, H., & Imuta, K. (2011). Episodic memory in 3- and 4-year-old children. *Developmental Psychobiology*, *53*, 317–322. https://doi.org/10.1002/dev.20527

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2007). *Empirical benchmarks for interpreting effect sizes in research*. New York, NY: MDRC.

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning & Verbal Behavior*, *11*, 534–537. https://doi.org/10.1016/S0022-5371(72)80036-7

Horn, D. L., Pisoni, D. B., & Miyamoto, R. T. (2006). Divergence of fine and gross motor skills in prelingually deaf children: Implications for cochlear implantation. *Laryngoscope*, *116*, 1500–1506. https://doi.org/10.1097/01.mlg.0000230404.84242.4c

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169669.003.0001

Icht, M., & Mama, Y. (2015). The production effect in memory: A prominent mnemonic in children. *Journal of Child Language*, *42*, 1102–1124. https://doi.org/10.1017/S0305000914000713

Icht, M., Mama, Y., & Algom, D. (2014). The production effect in memory: Multiple species of distinctiveness. *Frontiers in Psychology: Cognition*, *5*, Article No. 00886. https://doi.org/10.3389/fpsyg.2014 .00886

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*, 577–581. https://doi.org/10.3758/BF03214352

Jones, A. C., & Pyc, M. A. (2014). The production effect: Costs and benefits in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 300–305. https://doi.org/10.1037/a0033337

Jonker, T. R., Levene, M., & MacLeod, C. M. (2014). Testing the item-order account of design effects using the production effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 441–448. https://doi.org/10.1037/a0034977

Lin, O. Y.-H., & MacLeod, C. M. (2012). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology*, *66*, 212–216. https://doi.org/10.1037/a0028309

Litt, R. A., & Nation, K. (2014). The nature and specificity of paired associate learning deficits in children with dyslexia. *Journal of Memory and Language*, *71*, 71–88. https://doi.org/10.1016/j.jml.2013.10.005

Logan, J. A. R., Hart, S. A., Cutting, L., Deater-Deckard, K., Schatschneider, C., & Petrill, S. (2013). Reading development in young children: Genetic and environmental influences. *Child Development*, *84*, 2131–2144. https://doi.org/10.1111/cdev.12104

Louden, W., Chan, J. K. S., Elkins, J., Greaves, D., House, H., Milton, M., . . . van, Karryenoord, C. E. (2006). *Mapping the territory. Primary students with learning difficulties. Literacy and Numeracy* (Vol. 1–3). Canberra, ACT: Department of Education, Training and Youth Affairs.

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*, 390–395. https://doi.org/10.1177/0963721417691356

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 671–685. https://doi.org/10.1037/a0018785

Messer, D., & Dockrell, J. E. (2006). Children's word findings difficulties: Descriptions and explanations. *Journal of Speech, Hearing and Language Research*, *49*, 309–324. https://doi.org/10.1044/1092-4388(2006/025)

Michas, I., & Henry, L. A. (1994). The link between phonological memory and vocabulary acquisition. *British Journal of Developmental Psychology*, *12*, 147–163. https://doi.org/0.1111/j.2044-835X.1994.tb00625.x

Moghadam, S. H. (2012). A review on the important role of vocabulary knowledge in reading comprehension performance. *Procedia—Social and Behavioral Sciences*, *66*, 555–563. https://doi.org/10.1016/j.sbspro.2012.11.300

Mourgues, C., Tan, M., Hein, S., Ojanen, E., Reich, J., Lyytinen, H., & Grigorenko, E. L. (2016). Paired associate learning tasks and their contribution to reading skills. *Learning and Individual Differences*, *46*, 54–63. https://doi.org/10.1016/j.lindif.2014.12.003

Newcombe, N. S., Lloyd, M. E., & Ratliff, K. R. (2007). Development of episodic and autobiographical memory: A cognitive neuroscience perspective. *Advances in Child Development and Behavior*, *35*, 37–85. https://doi.org/10.1016/b978-0-12-009735-7.50007-4

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*, 326–338. https://doi.org/10.3758/s13421-011-0165-1

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory*, *20*, 717–727. https://doi.org/10.1080/09658211.2012.699070

Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory*, *22*, 509–524. https://doi.org/10.1080/09658211.2013.800554

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *Quarterly Journal of Experimental Psychology*, *44A*, 47–67. https://doi.org/10.1080/14640749208401283

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in Polyglots. *Quarterly Journal of Experimental PSychology*, *48*(1), 966–979. https://doi.org//10.1080/14640749508401378

Petersen, I. T., Bates, J. E., D'Onofrio, B. M., Coyne, C. A., Lansford, J. E., Dodge, K. A., , . . . Van Hulle, C. A. (2013). Language ability predicts the development of behavior problems in children. *Journal of Abnormal Psychology*, *122*, 542–557. https://doi.org/10.1037/a0031963

Putnam, A. L., Ozubko, J. D., MacLeod, C. M., & Roediger, III, H. L. (2014). The production effect in paired-associate learning: Benefits for item and associative information. *Memory & Cognition*, *42*, 409–420. https://doi.org/10.3758/s13421-013-0374-x

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory*, *21*, 904–915. https://doi.org/10.1080/09658211.2013.766754

Schacter, D. L., Israel, L., & Racine, C. A. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory & Language*, *40*, 1–24. https://doi.org/10.1006/jmla.1998.2611

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, *45A*, 21–50. https://doi.org/10.1080/14640749208401314

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign-language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, *16*, 155–172. https://doi.org/10.1017/S0142716400007062P

Share, D., & Silva, P. A. (1987). Language deficits and specific reading retardation: Cause or effect? *International Journal of Language & Communication Disorders*, *22*, 219–226. https://doi.org/10.3109/13682828709019864

Snowling, M. J., & Hulme, C. (2006). Language skills, learning to read and reading intervention. *London Review of Education*, *4*, 63–76. https://doi.org/10.1080/13603110600574462

Snowling, M., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Hulme, C. (2009). *York assessment of reading for comprehension: Passage reading [measurement instrument]*. London, UK: GL Assessment.

Stanovich, K. E. (1993). Dysrationalia: A new specific learning disability. *Journal of Learning Disabilities*, *26*, 501–516. https://doi.org/10.1177/002221949302600803

Stanovich, K. E. (2017). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education, Reading Research Quarterly*, *21*, 360–407. https://doi.org/10.1177/0022057409189001-204

Zamuner, T. S., Strahm, S., Morin-Lessard, E., & Page, M. P. A. (2017). Reverse production effect: Children recognize novel words better when they are heard rather than produced. *Developmental Science*, *21*, e12636. https://doi.org/10.1111/desc.12636

## Appendix A

### Experiment 1: Familiar Word Pool

| | | | | | |
|---|---|---|---|---|---|
| captain | century | forest | attention | foundation | debate |
| treasure | friend | peace | steam | history | record |
| office | lesson | direction | justice | ladder | turnip |
| knock | neighbor | handle | branch | daughter | travel |
| school | basket | whisper | winter | gravity | ocean |
| quarrel | invitation | uniform | shadow | teacher | invention |
| fashion | river | island | minute | account | author |
| leather | traffic | holiday | pocket | message | glass |
| package | afternoon | uncle | ticket | kitchen | garden |
| industry | avenue | journey | attitude | valley | orchard |
| resort | education | victory | trousers | envelope | quarter |
| election | amount | market | theater | wheel | wheat |
| shoulder | answer | meadow | language | stream | wagon |
| furniture | beauty | dinner | castle | queen | powder |
| partner | merchant | entrance | machine | address | arrow |
| border | evening | village | summer | porch | vacation |
| battery | pebble | judge | capital | distance | nephew |
| campaign | painting | factory | laugh | thread | engine |
| kettle | plate | clothes | reward | kingdom | speech |
| department | station | building | sailor | harbor | guardian |

## Appendix B

### Experiment 2: Novel Non-Word Pool

| | | | | | |
|---|---|---|---|---|---|
| beld | jame | pise | baunt | lorse | spoot |
| boke | jate | pote | binch | pench | stame |
| coof | kall | rark | bloss | porse | stell |
| cown | kend | rell | bouth | pouse | stope |
| coys | kurp | sark | chank | pribe | stort |
| dard | lafe | shog | chone | pross | trass |
| dirs | lecs | sife | clane | prown | trine |
| doke | losh | tarm | crace | routh | zight |
| dort | meam | teaf | creat | shace | blinch |
| fape | meck | turl | crove | shage | clunch |
| feen | mert | vate | datch | shalk | dridge |
| foat | mook | vead | dight | shart | prease |
| fuff | mout | vorn | drass | shork | prench |
| geal | nass | wase | drave | shunk | scarch |
| goss | neek | weam | fatch | slank | shrife |
| gowl | noke | wime | flass | slare | sprine |
| hean | noot | yeal | flink | slass | sprong |
| hest | pame | yide | grafe | slint | strofe |
| hoil | parl | zark | gress | spave | thrine |
| hout | pash | zear | jatch | spile | yought |