Received: 31 May 2016 | Accepted: 31 August 2017

DOI: 10.1111/desc.12636

PAPER

WILEY Developmental Science

Reverse production effect: children recognize novel words better when they are heard rather than produced

Tania S. Zamuner[1] | Stephanie Strahm[1] | Elizabeth Morin-Lessard[2] | Michael P.A. Page[3]

[1]Department of Linguistics, University of Ottawa, Ottawa, Canada

[2]Department of Psychology, Concordia University, Montreal, Quebec, Canada

[3]Department of Psychology, University of Hertfordshire, Hatfield, Hertfordshire, UK

Correspondence
Tania S. Zamuner, Department of Linguistics, University of Ottawa, Arts Hall, 70 Laurier Ave. East, Ottawa ON, Canada, K1N 6N5.
Email: tzamuner@uottawa.ca

Funding information
SSHRC grant awarded to TSZ and an FQRSC Doctoral Fellowship awarded to EML

Abstract

This research investigates the effect of production on 4.5- to 6-year-old children's recognition of newly learned words. In Experiment 1, children were taught four novel words in a produced or heard training condition during a brief training phase. In Experiment 2, children were taught eight novel words, and this time training condition was in a blocked design. Immediately after training, children were tested on their recognition of the trained novel words using a preferential looking paradigm. In both experiments, children recognized novel words that were produced and heard during training, but demonstrated better recognition for items that were heard. These findings are opposite to previous results reported in the literature with adults and children. Our results show that benefits of speech production for word learning are dependent on factors such as task complexity and the developmental stage of the learner.

RESEARCH HIGHLIGHTS

- We investigate the effect of production on children's word learning.
- Children were taught novel words that were either heard or produced during a brief training phase.
- Children had better recognition and recall for novel words that were heard rather than produced during training.
- The effects of production are dependent on task complexity and the developmental stage of the learner.

1 | INTRODUCTION

Infants have substantial knowledge about the lexical, phonological and grammatical structures of their language before they begin producing speech (Bergelson & Swingley, 2012; Curtin, Byers-Heinlein, & Werker, 2011; Curtin & Zamuner, 2014; Werker & Curtin, 2005). Although production is not a necessary component for word recognition, it would be hasty to assume that the mechanism and representations for word recognition are the same in preverbal infants as compared to children and adults who are able to both perceive and produce language. A central component of language communication involves the development of speech production skills, which entails knowledge about how sounds are articulated, the integration of sensory-motor cues, and the retrieval of stored representations from memory (McAllister Byun, Inkelas, & Rose, 2016; McAllister Byun & Tessier, 2016; Stoel-Gammon, 2011). As such, a major milestone in development is when children produce their first words. Moreover, young children will often spontaneously imitate new words, e.g., "Mother: They're called *eels*, Violet aged 2 years: *Eels*" (Providence Corpus; Demuth, Culbertson, & Alter, 2006). Children's imitation rates of new words, between the ages of 2 to 3 years, range between 17% to 43% (Zamuner & Thiessen, in press) to as high as 25% to 78% (Clark, 2007). The current research investigates the effect of speech production on lexical acquisition. Specifically, this research investigates whether children show better learning for novel words that were produced versus heard during a brief training phase.

1.1 | Converging evidence for a role of production in language development

Research on developmental speech production takes a somewhat indirect approach to examine the role or effect of production on learning—in other words, production is typically not a controlled factor. Evidence is emerging that production-based representations exist even before infants produce meaningful speech (Bruderer, Danielson, Kandhadai, & Werker, 2015; DePaolis, Vihman, &

Keren-Portnoy, 2011; DePaolis, Vihman, & Nakai, 2013; Majorano, Vihman, & DePaolis, 2014; Ngon & Peperkamp, 2016; Yeung & Werker, 2013). In addition to observing that there is a relationship between infants' babbling repertories and their first words (Vihman, Macken, Miller, Simmons & Miller, 1985), children are more likely to add words to their productive vocabulary when the words are shorter in word length, have more phonological neighbours (i.e., words that sound similar to many other words), and are more frequent (Carlson, Sonderegger, & Bane, 2014; Coady & Aslin, 2004; Maekawa & Storkel, 2006; Ota & Green, 2013; Stoel-Gammon, 1998; Storkel, 2004, 2006). Some work has also indicated a relationship between children's speech production skills and their lexical acquisition. Maekawa and Storkel (2006) found that in the data for one child out of three analysed from CHILDES, a significant predictor for whether a word was added to the child's vocabulary was the age of productive acquisition for words' final consonants. Zamuner and Thiessen (in press) examined data collapsed across five children aged 0;11–2;11 from the CHILDES Providence Corpus (Demuth et al., 2006), and found that a significant predictor for whether a child spontaneously imitated a new word was whether the word contained more consonants that the child had accurately produced before.

Other evidence for the role of production in language learning comes from the lexical avoidance effect, first noted in Ferguson and Farwell (1975). Although not universal, some children avoid producing or imitating words that contain sounds outside of their production repertoire; although this could be described as failing to select certain lexical items rather than avoidance (Schwartz & Leonard, 1982). Lexical avoidance has been indirectly examined in longitudinal studies (see studies described above) and experimentally. In a series of studies with children under 2 years of age (Leonard, Schwartz, Morris, & Chapman, 1981; Schwartz & Leonard, 1982), children's productive inventories were first assessed to establish which sounds were consistently produced (IN sounds) and what sounds were not attempted (OUT sounds). Children then were taught stimuli composed of IN sounds and OUT sounds, after which their learning was assessed using different measures, such as whether the child produced the word. Based on the data obtained from these measures, more words were learned when they contained IN sounds than OUT sounds, suggesting that children select or avoid certain words over others. The effect of previous production experience is also reported by Keren-Portnoy, Vihman, DePaolis, Whitaker, and Williams (2010), who found that children's non-word repetition is linked to their previous production experience. They tested the accuracy of 26-month-old children's non-word repetition, controlled for sounds found frequently (IN) or infrequently (OUT) in an individual child's previous production repertoire. Children were better at producing non-words comprising consonants frequently found in their production inventories. In work with older children aged 3 to 5 years, Storkel (2004, 2006) found a reversal of the effect of IN and OUT sounds on word learning, such that items which contained OUT sounds were learned better than items containing IN sounds. This indicates a possible developmental change in how existing phonological and lexical knowledge may guide lexical development. Yet, controlled studies are still needed that focus on the effect of production, independent from other variables (Core, 2012). For example, Zamuner, Fikkert, and Gick (2007) found that Dutch-learning children between 2 and 3 years of age who spontaneously produced novel words during a word-learning task were significantly better at naming at test; however, production during training was not a controlled variable. As such, production might not have been the only or primary factor in children's recall, as children who produced the novel words may have had more advanced language skills.

Combined, findings from a variety of paradigms provide support for the hypothesis that production positively impacts language learning. Most of this research comes from studies of infants and young children, who are at the beginning stages of learning how to produce speech. The results from this research are consistent with an "articulatory filter" (Vihman, 1993, Vihman, DePaolis, & Keren-Portnoy, 2014), which posits that previously produced sounds are more salient in the learner's input. Under this account, words containing previously produced sounds are also theorized to require less processing resources during word learning, allowing for more resources for the mapping of a word form to a referent. Additional support for this is seen in experimental studies with older children and adults, in which an initial speech practice phase improves participants' performance on a fast mapping task (Kan & Sadagopan, 2015; Kan, Sadagopan, Janich, & Andrade, 2014), and from work showing that children showed better memory for words' plurality when the words contained sounds within children's production repertoire (Ettlinger, Lanter, & Van Pay, 2014). These results are consistent with exemplar theories and usage-based models proposing that speech output representations are integrated into phonological and lexical representations (Munson, Edwards, & Beckman, 2012; Pierrehumbert, 2003; Sosa & Bybee, 2008; Sosa & Stoel-Gammon, 2012), and also in the recently developed A-map model of developmental speech production (McAllister Byun et al., 2016). Despite these accounts, notably absent in the literature are studies that systematically control production during first language acquisition (although see work looking at the effect of lexical status on the stability of children's articulatory movement trajectories using speech kinematic measurements; Heisler, Goffman, & Younger, 2010). We now turn to a discussion of controlled studies of production with adult participants and to a study done with children.

1.2 | The production effect

MacLeod, Gopie, Hourihan, Neary, and Ozubko (2010) termed the *production effect* which is "the fact that producing a word aloud during study, relative to simply reading a word silently, improves explicit memory" (MacLeod et al., 2010, p. 671). In a series of studies with adults built on previous findings (Conway & Gathercole, 1987; Gathercole & Conway, 1988; Hopkins & Edwards, 1972), MacLeod et al. (2010) found that words read aloud or mouthed silently were recognized better than words which were read silently. This is accounted for by appealing to distinctiveness, where "the act of speaking those words aloud is encoded and later recovery of this information can be used to infer that those words were studied" (Ozubko, Major, & MacLeod,

2014, p. 509). Since their initial study in 2010, multiple studies have reported similar findings with adults in behavioral research (Fawcett, 2013; Kaushanskaya & Yoo, 2011; Mama & Icht, 2016; MacLeod, 2011; Ozubko & MacLeod, 2010; Zamuner, Morin-Lessard, Strahm, & Page, 2016, among others) and neurophysiological research (Mathias, Palmer, Perrin, & Tillmann, 2015). The production effect is strongest in within-subject designs (Fawcett, 2013), and there are studies showing no effect of production in between-subjects designs (e.g., Abbs, Gupta, & Khetarpal, 2008).

In addition to appealing to distinctiveness, researchers have suggested that production strengthens or enriches representations. Zamuner et al. (2016) suggest that production includes additional articulatory phonetic information, leading to faster word recognition. This account is consistent with a variety of models that incorporate articulatory representations and/or sensory-motor components, such as exemplar theories (Goldinger, 1998; Pierrehumbert, 2003; Sosa & Bybee, 2008) and more recent neurocognitive models of speech production (Dell, Schwartz, Nozari, Faseyitan, & Coslett, 2013; Hickok, 2012). The findings are also consistent with models that relate Hebb repetition learning to word learning (Gupta & Tisdale, 2009; Page & Norris, 2009). Producing an item during learning may create a bidirectional link between production and comprehension, resulting in improved recognition for items produced over items heard during training. Production effects have also been accounted for by the phonological loop, a component of Baddeley and Hitch's (1974) model of working memory that supports the processing of phonological material and word learning (Baddeley, Gathercole, & Papagno, 1998; Gathercole, 2006; Gathercole, Pickering, Ambridge, & Wearing, 2004). The phonological loop has two parts: a phonological store (storing sequences of speech sounds) and an articulatory rehearsal component (repeating or looping material to maintain it in memory). Overt vocal rehearsal is hypothesized to create stronger links with long-term phonological knowledge, leading to better learning (Gathercole & Conway, 1988; Kaushanskaya & Yoo, 2011).

To date, only one study has looked at the production effect with children using a similar design as in the adult literature (Icht & Mama, 2015). In their first experiment, 5-year-old Hebrew-speaking children had the highest recall rates for familiar words that were produced during training (look and say), followed by words that were heard (look and listen—produced by the experimenter), followed by silent observation (look). In a follow-up study, they extended their findings to unfamiliar words. Unfamiliar was defined as words not in children's expressive vocabulary, for example, anchor, manger, cuff, pestle, and razor. The training phase had two conditions: look-and-say and look-and-listen. In a four-alternative forced-choice test, 5-year-old children saw the target, a distractor learned through the same condition as the target, a distractor learned through the other condition, and a distractor not present in the training phase. The experimenter produced the target word and children were asked to point to the appropriate image. Children recognized more words that were produced than heard during training. These findings were accounted for by appealing to distinctiveness and the number of unique processes in which the words were encoded. They argued that there was an advantage for produced items (look and say) because their encoding involved three processes: visual, "articulation (the execution of a motor action) and audition (hearing the word)" (Icht & Mama, 2015, p. 1103), whereas the other conditions (look and listen, look) had fewer encoding processes.

Yet production does not always lead to improved recall and recognition of items. A handful of studies with adults have reported that in certain circumstances, the production effect can be attenuated or reversed (Cho & Feldman, 2013, 2016; Dahlen & Caldwell-Harris, 2013; Kaushanskaya & Yoo, 2011; Zamuner, Strahm, Morin-Lessard, & Page, submitted). When trained on novel words with native phonemes in either an overt vocal training condition or a subvocal (silent rehearsal) training condition, adults showed better recall and recognition in the overt production condition. However, this effect was reversed when the novel words contained non-native phonemes (Kaushanskaya & Yoo, 2011). Similarly, when adults in Zamuner et al. (submitted) were trained on novel words with high phonotactic probabilities, a production advantage was found in recall. But when the novel words comprised low phonotactic probabilities, the effect was attenuated and there was no production advantage in recall. Baese-Berk and Samuel (2016) found that production disrupted adults' learning of a novel fricative contrast compared to a heard condition. In Kaushanskaya and Yoo (2011) and Baese-Berk and Samuel (2016), adults did not always produce target-like productions when they repeated the stimuli aloud. Therefore, part of the decreased learning in the overt production condition may have stemmed from the mismatch between the auditory target and participants' non-target-like productions. Stimuli with non-native phonemes and with infrequent sound patterns may also require more processing resources than stimuli with native and frequent sound patterns. Interestingly, Baese-Berk and Samuel also found a reduction in learning of a novel fricative contrast when the production system was engaged in another task—participants named letters on the screen during training. Baese-Berk and Samuel suggest that part of the disruption in learning may also stem from "cognitive mechanisms such as selective attention and task switching" (Baese-Berk & Samuel, 2016, p. 31).

Together, these studies lead to the important observation that the direction and/or strength of the production effect may depend on the stimuli and the difficulty of the task (Zamuner, Yeung, & Ducos, 2017). Perceptual salience, task demands, and the developmental stage of the learner are modeled in the PRIMIR framework (Processing Rich Information from Multidimensional Interactive Representations; Curtin et al., 2011; Werker & Curtin, 2005). In PRIMIR, task demands interact with the developmental level of the learner with respect to what representations are accessed and processed. (Also see research on lexical input processing that considers how attention and processing resources are allocated during second language learning, summarized in Barcroft, 2015.) While PRIMIR is a model of developmental speech perception, emerging speech production skills can be extended within their model. When a child is completing a difficult task, such as producing new words, the mapping between the form and the referent may be disrupted. Potential tasks effects are relevant to the current studies, because rather than looking at the effects of production on longitudinal and spontaneous corpora (e.g., Maekawa & Storkl, 2006;

Vihman et al., 1985; Zamuner & Thiessen, in press), we experimentally manipulate children's production (or not) of new words.

1.3 | The current research

To this point, we have discussed effects of production on language learning, yet learning can be defined and measured in numerous ways, and it is an open question as to how production may impact different levels of representations and processing, in both short-term and long-term learning. In addition, it is important to keep in mind that the impact of production may vary across development, with different effects of production for a child who is just learning how to speak, compared to an older child who is proficient in comprehending and producing language. The goal of this research was to examine whether production affects lexical acquisition, as measured by novel word recognition. Many of the previous methodologies used with adults are not suitable for young children, especially tasks that involve reading. We designed an experiment in which participants were trained on a set of auditory presented novel words with their visual referents. While our long-term goal is to develop a methodology that is suitable to test the effects of production with toddlers, our initial research is with 4.5- to 6-year-old children. This allows us to establish a working methodology and enables us to compare the effect of production to a similar aged group of children already studied by Icht and Mama (2015). During a brief training phase, children were taught novel words and asked to either repeat the novel words or to be silent. Immediately after training, children were tested on their recognition of the novel words using a preferential looking paradigm. This same task was used in a study with adult participants (Zamuner et al., 2016), who showed better recognition for novel words produced over those heard during training. We predicted that if production enhances children's learning in a similar way, that the same effect would be found. Children's recognition accuracy would be higher for the novel words that were produced compared to those heard during training. However, if the task was challenging for children, as modelled in developmental frameworks (Curtin et al., 2011; Vihman et al., 2014; Werker & Curtin, 2005), we predicted that the production effect would be attenuated or reversed.

2 | EXPERIMENT 1

2.1 | Method

2.1.1 | Participants

Participants were 16 English-speaking children (12 males, 4 females), between the ages of 4.5 and 6 years (*M* = 5;5, range = 4;5 to 6;8). Participants were required to be English dominant, with no more than 25% exposure to another language. Language background and exposure was determined by parental language questionnaire. Children were reported having exposure to a variety of languages, the most common was French, but some reported exposure to Chinese, Italian, Dutch, Arabic or German. All children were also reported to have normal hearing, normal vision, and no history of language impairment. We used a questionnaire rather than standardized tests because of time constraints inherent to one of our testing locations. Children were tested in one session, at a university campus or in a museum-based lab—the latter of which were limited to 15-minute sessions. Eleven additional participants were tested but not included in the analysis because they had no video for off-line coding (2), had no data in a condition (1), or had an error on one or more of the training trials: produced a heard item (7), mispronounced a target word (1).

2.1.2 | Stimuli

The stimuli consisted of two sets of four CVC novel words: Set 1 /wɪs, zɛl, gʌb, mig/, Set 2 /nɪs, kɛl, fʌb, jig/. Participants were tested either on Set 1 or Set 2, with counterbalancing: List 1: Heard /wɪs, zɛl/, Produced /gʌb, mig/, List 2: Heard /gʌb, mig/, Produced /wɪs, zɛl/, List 3: Heard /nɪs, kɛl/, Produced /fʌb, jig/, List 4: Heard /fʌb, jig/, Produced /nɪs, kɛl/. All stimuli were pre-recorded by a native speaker of English, and normalized for amplitude (70 dB). Stimuli were not controlled for their phonotactic probabilities, given the need to have novel words with as many unique consonants and vowels as possible, to avoid overlap between the items. However, the final set of stimuli were later checked to ensure that the novel words did not comprise low frequency sound patterns. This was done based on the standards in the field of calculating the stimuli's segmental positional frequencies (frequency of occurrence for individual consonants and vowels) and biphone frequencies (frequency of co-occurrence for consonant-vowel and vowel-consonant). The calculations were based on a corpus of child speech (Storkel & Hoover, 2010), and are provided in supplementary materials in Storkel (2013). The stimuli's characteristics were as follows: positional segmental sum, *M* = 0.15, *SD* = 0.05; biphone sum, *M* = 0.0075, *SD* = 0.005. These numbers are higher than averages for all English CVC non-words (see Table 1,

TABLE 1 Experiment 1 GCA results. Empirical logit GCA for the effect of Training on looking data for a window of analysis 200–1500 ms after the target novel word onset. The values in each cell represent parameter estimates

Predictor	Estimate	*SE*	*t*	*p*
Intercept	1.52	0.31	4.85	<.001
Training (produced vs. heard)	−1.04	0.47	−2.23	.03
Linear term	2.23	0.76	2.93	.003
Quadratic term	−0.66	0.80	−0.82	.41
Training * Linear term	0.14	0.84	0.16	.87
Training * Quadratic term	0.54	0.84	0.64	.52

Note. *SE* = standard error. GCA structure in R: lmer(elog ~ Training * (Linear + Quadratic) + (1 + Training + Linear + Quadratic | Participants). *p* = *p*-values, calculated in R using a normal approximation, based on the assumption that the *t*-distribution converges to the *z*-distribution.

Storkel, 2013). The stimuli also fall within the mid-high to highest phonotactic probability quartiles reported for the stimuli in Storkel, Bontempo, Aschenbrenner, Maekawa, and Lee (2013). The neighborhood density of the stimuli was also calculated based on the same corpus and were as follows: range = 1 to 16, *M* = 7.5, *SD* = 4.92. The neighborhood density of the items was spread across the range of low to highest neighbourhood density quartiles reported for the stimuli in Storkel et al. (2013). The visual stimuli consisted of four nonce animals (Ohala, 1999). Training condition and image association of novel words were counterbalanced across four lists.

2.1.3 | Design

The design was a shorter version of an adult study by Zamuner et al. (2016). Participants first completed a practice study which familiarized children to the task. The practice study had the same structure as the experiment, but with fewer trials (six practice, six test) and with familiar words (e.g., banana, apple). The practice study was followed by the experiment, which consisted of eight training trials, followed by eight test trials (Figure 1). During training, two novel words were assigned to the produced condition (two training trials for each novel word = 4 trials) and two novel words were assigned to the heard condition (two training trials for each novel word = 4 trials). Participants heard an auditory token of a novel word with the corresponding nonce animal image. After 2000 ms, a prompt image appeared below the nonce animal. If the picture was a finger pointing towards the participant, they were instructed to repeat the novel words. If the picture was of a woman gesturing "shh" with her finger over her lips, they were instructed to remain silent, and a second presentation of the same audio token was presented 500 ms after the prompt image. This meant that participants did not know whether the novel word was to be produced or heard until after the prompt image appeared on the screen. This design also provided a control for frequency during training because production increases an item's frequency (Abbs et al., 2008; Hopkins & Edwards, 1972): produced items were presented twice during training (once from the computer and once by the participants), and heard items were also presented twice (two times by the computer). A similar control for the frequency of presentation was also done in two studies with adults (Zamuner et al., submitted; Zamuner et al., 2016), and in a study with 5-year-olds by Icht and Mama (2015) which used a live experimenter, and on heard trials the items were produced twice by the experimenter.

At test, two images were presented for 2000 ms in silence, then a target novel word was named (the same audio token as used during training). Then 1500 ms later, participants heard "Do you see it?" or "Do you like it?" to maintain their attention to the screen. Trials were separated with a blank screen for 2000 ms. Images were paired so that both were produced or both heard during training. This was to avoid the possibility that a participant might recognize an item from one condition more because it was more salient relative to the other condition. Each novel word was used as the target twice (eight test trials) and counterbalanced for which side of the screen the target appeared on (left or right). Images were horizontally centered, spaced 420 pixels apart and sized 400 × 400 pixels. An Eyelink 1000 Remote eyetracker tracked the eye gaze of participants using monocular tracking. The eyetracker has a sampling rate of 500 Hz (SR Research, Ottawa), and

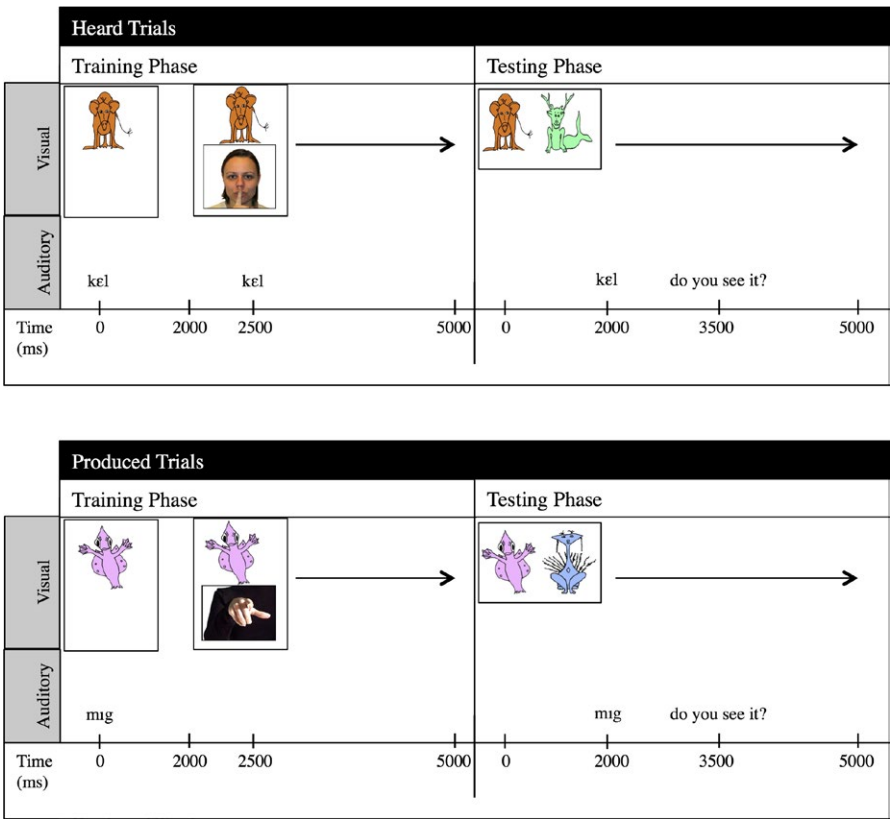

FIGURE 1 Time course of heard and produced training and test trials

was mounted on a 17-inch monitor displaying the stimuli. Experiment Builder software was used for data collection.

2.1.4 | Procedure

Children were individually tested in a sound-attenuated booth at a university campus or on the same eyetracker that was transported to use in a sound-attenuated room at a local museum. The eyetracker was calibrated using a 5-point grid arranged in a "plus sign" shape, using a bird's face. Drift correction was performed between every test trial in the form of a central fixation image to account for shifts in eye position. In addition to the eyetracking data, the experiment was recorded with an external video camera to allow for coding of potential mistakes during the training trials. These training phase videos were coded off-line for the type of response (produced or heard). There were only two novel words for each condition; therefore, if a child made a mistake during training such as producing a novel word in the heard condition, that incorrectly trained novel word would be either a target or a distractor for the heard condition at test. Seven participants were excluded for producing a heard item during training. The training trials were also transcribed by a researcher trained in phonetics, and coded for the accuracy of the produced response (correctly pronounced or incorrectly pronounced), for example /zɛl/ produced as /tzɛl/. One participant was excluded for making a mispronunciation error during training (note that a subset of the seven participants that were excluded for producing items on heard trials also made mispronunciation errors). Participants were tested in one session, lasting approximately 10 min.

2.1.5 | Analyses

Areas of interest around the images were created using the Experiment Builder software, and were the same size as the images, 400 × 400 pixels. All looking outside of these areas was excluded from the analyses. For each trial, the proportion of looking to the target image (vs. the distractor) was split into 100 ms time bins. The window of analysis started 200 ms after the beginning of the target, accounting for the time it takes to initiate eye movements (Matin, Shao, & Boff, 1993; Salverda, Kleinschmidt, & Tanenhaus, 2014). Note that in studies with children, analyses can also begin 300 ms after the target, due to longer latencies in children's eye movements (Buckler & Fikkert, 2016; also see Swingley & Aslin, 2000, for a discussion of latencies). All analyses presented below were also done starting at 300 ms after word onset, with the same results. Children's fixation to the target peaked at 1500 ms after target word onset. This resulted in a window of analysis from 200 to 1500 ms, with 14 time bins in total. Participants' proportional looking to the target was analyzed using a growth curve analysis (GCA) performed in R (R Core Team) using the lmer() function from the lmer4 package (version 1.1-7; Bates, Maechler, Bolker, & Walker, 2014). GCA is a multilevel regression method allowing for the assessment of both the differences in time-spent-looking and in the steepness of a looking curve that is taken to indicate lexical access (Mirman, Dixon, & Magnuson, 2008). The empirical logit was calculated for each time bin as an approximation to log odds (Barr, 2008).

2.2 | Results

2.2.1 | Training trials

Analyses first compared total looking during the different training conditions to assess whether children looked longer on the produced or heard training trials. A two-tailed paired-samples *t* test revealed no statistically significant difference in the mean total looking time on produced training trials (M = 1359 ms) versus heard training trials (M = 1416 ms), $t(15) = 0.43$, $p > .05$. There was also no statistically significant difference in looking times after the appearance of the prompt images for produced training trials (M = 679 ms) versus heard training trials (M = 708 ms), $t(15) = 0.80$, $p > .05$. These analyses rule out total looking time during training as a potential influence in participants' looking to targets on produced and heard test trials.

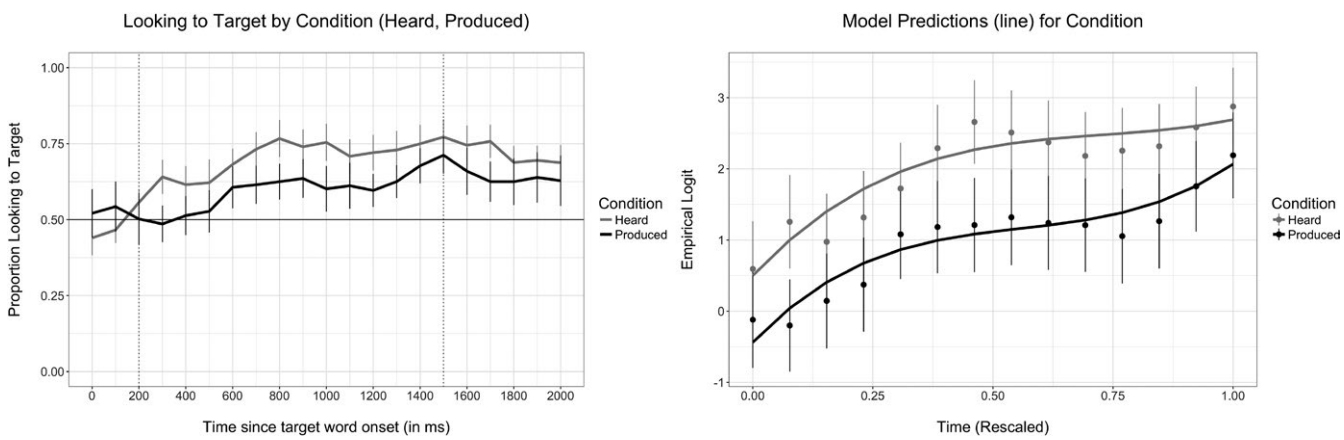

FIGURE 2 Experiment 1 proportion of looking graph and model predictions. (Left) Time course of proportion of looking to the target starting after non-word onset for novel word from produced and heard training conditions. The horizontal dotted line corresponds to chance (0.50), the vertical line throughout each point represents the standard error of the mean, and the dotted vertical lines indicate the window of analysis. (Right) Model predictions for Training for 200–1500 ms window of analysis based on proportions of looking to target. Points show actual values

2.2.2 | Test trials

Figure 2 provides the proportion of looks to the target as a function of time in the two conditions, spanning from word onset (0 ms) to 2 secs afterwards. The model contained Training (produced, heard, a within-subjects variable) as the main predictor, which was deviation-coded. The model also included main effects of Training and Time (modeled with second-order orthogonal polynomials) and a Training/Time interaction. Random by-participants effects were included for the intercepts and random slopes on the time terms (linear and quadratic). See Table 1 for the full model. There was a statistically significant effect of training (produced, heard) overall (Estimate = −1.04, *SE* = 0.47, *p* = .03), but no statistically significant effect of training on the linear term (Estimate = 0.14, *SE* = 0.84, *p* = .87) or quadratic term (Estimate = 0.54, *SE* = 0.84, *p* = .52). This indicates that there was more looking to targets on heard versus produced trials, but there was no statistically significant difference in the slope or shape of the looking curves between the conditions (Figure 2).[1] We then compared participants looking to the target compared to a chance level of 50% fixation to one of the two images, at each 100 ms time bin across the 200 to 1500 ms window after target word onset. Looking to the target on the heard trials differed from chance starting 600 ms after the target word onset, and the greater looking compared to chance extended throughout the rest of the window (based on sliding two-tailed *t* tests with an alpha adjustment to .0036 based on 14 multiple comparisons, one for each time bin), reaching the maximum of 76% looking to heard targets. In contrast, looking to the target on the produced trials was only statistically significantly different at 1500 ms after the target word onset (alpha level .0036), and the maximum looking to produced targets was 71%.

2.3 | Discussion

In Experiment 1, children successfully mapped novel words in both the produced and heard training condition, as indicated by greater looking to the target compared to chance at test. Overall, children also looked more to targets in the heard condition, suggesting that children learned the novel words better in the heard than in the produced condition. However, the slope and the shape of the looking curves were not different in the two conditions. Differences in the overall looking to heard targets at test could not be attributed to differences in exposure during training, as children looked equally long to the screen during produced and heard training trials.

Our results were unexpected given the majority of studies in the literature that report a benefit for production. Especially given that when the same experiment was run with adults, a recognition advantage was found for the novel words that are produced over heard during training (Zamuner et al., 2016). Despite this, our findings are in line with a subset of studies that have found that the production effect can be attenuated or reversed, depending on the linguistic characteristics of the stimuli and potentially the difficulty of the task. The reversal of the production effect in our study cannot be attributed to non-native or infrequent sound patterns of the stimuli because the novel words in Experiment 1 were all phonotactically legal, and fell within the mid-to-high phonotactic probability ranges for children (Storkel et al., 2013). Moreover, disruptions in learning were not caused by mismatches between the auditory target and children's non-target productions because all of the stimuli were correctly produced by the children included in the analyses. Another possibility is that differences in the variability of the training stimuli might have led to differences in mapping. In the produced condition, items were presented once by the computer and then once through children's own productions, whereas heard items were presented twice by the computer. This cannot fully explain the direction of the effect though because the classic production effect was found when the same experiment was run with adult participants (Zamuner et al., 2016). However, the issue of variability could be addressed in future research by manipulating the amount of variability presented to participants in the different conditions.

We propose that the reversal of the production effect more likely stems from the difficulty of the task, from multiple sources. The complexity of our task might partly stem from the use of novel words and nonce referents, both of which have been shown to increase the processing load in word recognition tasks in studies with 14-month-old infants (Fennell, 2012; Fennell & Werker, 2003). While Icht and Mama (2015) found a benefit in recognition for produced items over heard items with low frequency nouns, they defined low frequency as words not likely to be found in children's expressive vocabulary (e.g., anchor, manger, cuff, pestle, razor). As such, children likely had some receptive knowledge of the infrequent word forms and their visual referents. The complexity of our task might also stem from our visual stimuli which consisted of all nonce animals. Even though our nonce animals were chosen to be maximally distinct in terms of color and features, all our items were animate, and consisted of transposing and mismatching existing features from animals, such as the body of a fish morphed onto a head with antlers. Another possibility is that task difficulties stemmed from a combination of intermixing of produced and heard training trials, where the appropriate response for participants was not indicated until the response prompt appeared. While the intermixing of produced and heard training trials was also a feature of Icht and Mama's study, the appropriate response cue for children was indicated throughout the trials.

We identified a number of possible sources which may have made our task cognitively demanding. In Experiment 2, we sought to determine whether children would show a different pattern of learning during produced and heard training trials if the task was simplified. The predictions for Experiment 2 were in line with those made for Experiment 1. If production enhances the representations that children have for newly-learned words, their recognition should be improved for novel words that were produced rather than heard during training, but if production makes the recognition task too challenging for children, the expected production effect should be diminished or reversed.

3 | EXPERIMENT 2

As in Experiment 1, children in Experiment 2 were trained on novel words that were produced or heard during training, and at test their

learning was assessed using a recognition task. The design was the same as in Experiment 1; however, the different training conditions were blocked rather than combined. It was expected that this would simplify the task because participants would not need to switch between produced and heard trials during training. In Experiment 1, participants learned four novel words (two from each training condition). A decision was made to increase this number to eight novel words in Experiment 2 (four from each training condition) to avoid potential ceiling effects resulting from children having to learn only two novel words in each blocked training condition. Lastly, an additional recall task was added to assess whether the production effect would be found when children had to recall the novel words, compared to the recognition task.

3.1 | Method

3.1.1 | Participants

Participants were 16 English-speaking children (12 males, 4 females), between the ages of 5 and 6 years (*M* = 5;11, range 5;2 to 6;8). Participants were required to be English dominant, and to have no more than 25% exposure to another language (French). Eight additional participants were tested but not included in the analysis due to failing to complete the experiment (2), equipment error (1), no data in a condition (1) or had an error on one or more of the training trials: produced a heard item (2), did not produce a produced item (2). All children had normal hearing, normal vision, and no history of language impairment, as determined by parental questionnaire.

3.1.2 | Stimuli

The stimuli consisted of two sets of eight CVC novel words: Set 1 /wɪs, zɛl, gʌb, mig, vup, rem, dæs, bos/, Set 2 /nɪs, kɛl, fʌb, jig, tup, pem, hæs, los/. The stimuli were counterbalanced for training condition: List 1: Heard /wɪs, zɛl, vup, bos/, Produced /gʌb, mig, rem, dæs/, List 2: Heard /gʌb, mig, rem, dæs/, Produced /wɪs, zɛl, vup, bos/, List 3: Heard /nɪs, kɛl, tup, los/, Produced /fʌb, jig, pem, hæs/, List 4: Heard /fʌb, jig, pem, hæs/, Produced /nɪs, kɛl, tup, los/. The stimuli's characteristics were as follows: positional segmental sum, *M* = 0.14, *SD* = 0.05; biphone sum, *M* = 0.0075, *SD* = 0.005. These numbers are higher than the averages based on the Child Corpus (see Table 1, Storkel, 2013), and fall within the mid-high to highest phonotactic probability quartiles reported in Storkel et al. (2013). The neighborhood density of the items was as follows: range = 1 to 19, *M* = 19.06, *SD* = 5.42. The range of the items' neighborhood densities was spread across the range of low to highest quartiles reported for the stimuli in Storkel et al.

3.1.3 | Design

The design was the same as in Experiment 1, and only the differences will be outlined here. In Experiment 2, the two training conditions were blocked rather than intermixed. The order of the training condition blocks was counterbalanced. Before each block, participants first completed a practice block that had the same structure as the produced or heard block, but with fewer trials and with familiar words. During the experimental blocks, children were taught four novel words, followed by the corresponding test trials.

3.1.4 | Procedure

The procedure was the same as in Experiment 1. In addition, at the end of the study, participants also performed a recall task, and were asked "Do you remember any of the names of the animals?" This was to look at the effect of produced versus heard training on participants' recall, and to compare children's performance on a recognition task versus recall task.

3.1.5 | Analyses

These were the same as in Experiment 1.

3.2 | Results

3.2.1 | Training trials

There were no differences in overall looking during training between the produced training trials (*M* = 1402 ms) and heard training trials (*M* = 1323 ms), $t(15) = -0.37$, $p > .05$. There was also no statistically significant difference in looking times after the appearance of the prompt image between produced training trials (*M* = 716 ms) and heard training trials (*M* = 697 ms), $t(15) = -0.16$, $p > .05$.

3.2.2 | Test trials

Children's looking to the target was analyzed using a GCA, with the same time window of 200–1500 ms after novel word onset. See Table 2 for the full model. While there was more looking to targets on heard versus produced trials, there was no statistically significant effect of Training (produced, heard) overall (Estimate = −0.82, *SE* = 0.49, $p = .10$). There was a statistically significant effect of training on the linear term (Estimate = −1.88, *SE* = 0.66, $p = .004$), but not on the quadratic term (Estimate = −0.22, *SE* = 0.65, $p = .74$). Thus, there was a difference in the slope, but not the shape of the looking curves between the two conditions (Figure 3). The slope of heard targets was steeper than that for produced targets, indicating faster looking to the target in the heard condition.[2] Looking to the target on the heard trials differed from chance (50% fixation to the target image) starting 1100 ms after target word onset, and the greater looking compared to chance extended throughout the rest of the window (based on sliding two-tailed *t* tests with an alpha adjustment to .0036 based on 14 multiple comparisons, one for each time bin), reaching 83% fixation to heard targets. This is later than in Experiment 1 where looking to heard targets was statistically different from chance starting at 600 ms and peaked at 76% looking to heard targets. For the produced trials, looking to targets was not statistically significant different from

chance starting (alpha level .0036), and reached 67% looking to target. (Some time periods approached significance; e.g., at 1200 ms the *p*-value was .013.) This is partially similar to Experiment 1, where looking to produced targets was statistically different only 1500 ms after target word onset and where the maximal looking to produced targets was 71%. A final GCA analyses was done to compare the results from Experiment 1 and Experiment 2, using the same frame of 200 to 1500 ms after non-word onset. Fixations to target was modeled in the same way as in the previous analysis, with an additional variable of Experiment. The intercept was significant (Estimate = 1.43, *SE* = 0.24, $p < .001$), and there was a significant effect of Training (produced, heard) overall (Estimate = −0.93, *SE* = 0.34, $p = .007$), and a significant overall effect of training on the linear term (Estimate = 2.48, *SE* = 0.47, $p < .001$). There was no significant overall effect of Experiment (Experiment 1, Experiment 2) (Estimate = −0.18, *SE* = 0.49, $p = .72$).

TABLE 2 Experiment 2 GCA results. Empirical logit GCA for the effect of Training on looking data for a window of analysis 200–1500 ms after the target novel word onset. The values in each cell represent parameter estimates

Predictor	Estimate	*SE*	*t*	*p*
Intercept	1.34	0.37	3.63	<.001
Training (produced vs. heard)	−0.82	0.49	−1.67	.10
Linear term	2.74	0.55	4.97	<.001
Quadratic term	−0.19	0.35	−0.54	.59
Training * Linear term	−1.88	0.66	−2.87	.004
Training * Quadratic term	−0.22	0.65	−0.33	.74

Note. *SE* = standard error. GCA structure in R: lmer(elog ~ Training * (Linear + Quadratic) + (1 + Training + Linear + Quadratic | Participants). *p* = *p*-values, calculated in R as described under Table 1.

After both blocks were completed, children were asked if they remembered the names for any of the animals. Two children were not asked because they were no longer interested in participating in the task. Coding of children's recalled items was done in two ways. First, responses were coded based on whether the initial consonant and final consonant matched target items, ignoring the quality of the vowel. If a child recalled /mʌg/, it could be considered as a match to either /mig/ or /gʌb/, but in this analysis it was coded as match for /mig/ because of the overlap in both the initial and final consonant. A two-tailed *t* test on the number of produced and heard items indicated a statistically significant difference, with greater recall for heard items ($M = 1.07$, $SD = 1.0$) compared to produced items ($M = 0.21$, $SD = 0.58$), $t(13) = -2.78$, $p = .01$, $d = 1.16$. The second analyses only considered correctly recalled items, where all the consonants and vowels had to be correctly recalled for a response to be coded as correct. There were too little data to statistically analyze because only eight children correctly recalled items; however, the results were in the direction predicted based on the eyetracking data: nine heard items were recalled and one produced item was recalled.

3.3 | Discussion

Children in Experiment 2 successfully mapped targets in the heard condition, indicated by greater looking to the target compared to chance. Looking to the targets in the produced condition differed from chance, but did not go below the alpha level of .0036. The different pattern of looking to targets in the heard versus produced conditions could not be attributed to differences in the amount of looking to the screen during training. In addition, the slope of looking in the heard condition was steeper than in the produced condition, indicating faster looking to the target in the heard condition. Finally, in a free recall task, more items were recalled that had been heard during training than produced during training. The number of items that were recalled was quite low compared to children's recognition rates. Differences in performance on recognition and recall tasks have been reported

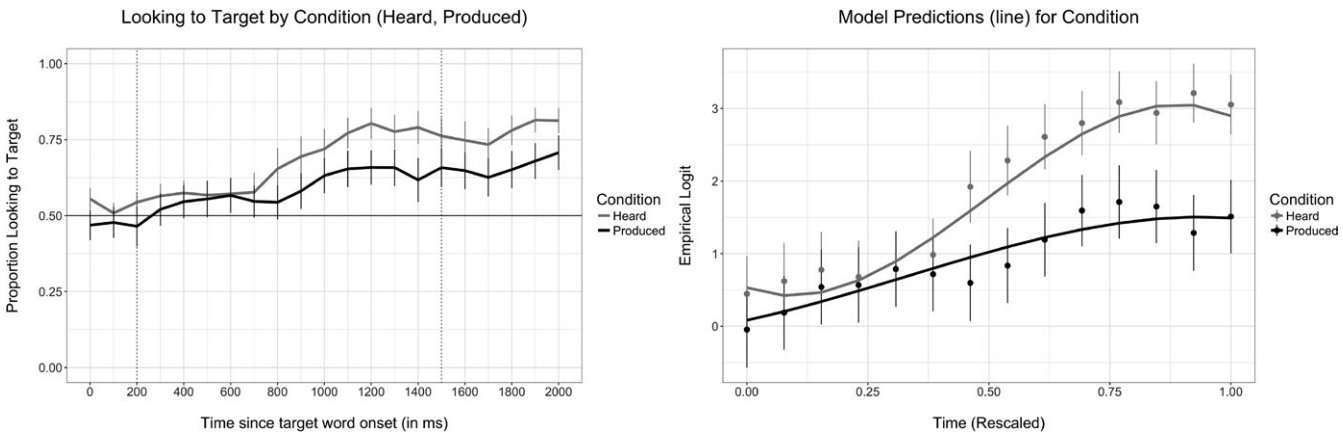

FIGURE 3 Experiment 2 proportion of looking graph and model predictions. (Left) Time course of proportion of looking to the target after non-word onset for novel word from produced and heard training conditions. The horizontal dotted line corresponds to chance (0.50), the vertical line throughout each point represents the standard error of the mean, and the dotted vertical lines indicate the window of analysis. (Right) Model predictions for Training for 200–1500 ms window of analysis based on proportions of looking to target. Points show actual values

before in the literature. In a study of fast mapping with 2-year-olds, Munro, Baker, McGregor, Docking, and Arciuli (2012) report that children had higher recognition rates compared to recall rates (also see Hodges, Munro, Baker, McGregor, Docking, & Arciuli, 2016).

4 | GENERAL DISCUSSION

This is one of the few studies with children that provide controlled data on the effects of production on word learning. In both Experiments 1 and 2, children successfully mapped novel words to their nonce referents. However, contrary to the majority of results reported from production studies with adults and the only existing research on the production effect with children (Icht & Mama, 2015), in both of the current studies, children showed better recognition for novel words that were heard rather than produced during the time of training. We call this *the reverse production effect*, as a counterpart to the previously coined term *the production effect* (MacLeod et al., 2010). While our findings may appear counter-intuitive given the previous literature, there are few controlled studies with children that directly ask this question. In addition, previous studies with adults have found that the production effect can change depending on the stimuli and task. Thus, our results are consistent with those previous findings.

Converging evidence from a variety of methodologies and ages has shown that in other circumstances, production facilitates learning; thus, it is a thought-provoking challenge to incorporate the current findings into existing theories of language development. We build as a starting point from the existing work with adults that has found that the production effect can be attenuated or reversed when the linguistic stimuli are non-native or infrequent (Baese-Berk & Samuel, 2016; Dahlen & Caldwell-Harris, 2013; Kaushanskaya & Yoo, 2011; Zamuner et al., submitted) and that learning can be disrupted when the production system is engaged in another task (Baese-Berk & Samuel, 2016). With both our mixed training conditions (Experiment 1) and blocked training conditions (Experiment 2), participants showed better mapping in the heard condition than in the produced condition. What then about production is more cognitively demanding in our task? We hypothesize that it is cognitively demanding for children to simultaneously map a novel word form and its novel referent, while simultaneously engaging the production system in repeating the novel word (also see Munro et al., 2012, for a discussion on cognitive resources during fast mapping). In the heard condition, participants may have more processing resources to map the word form and the referent because the production system is not engaged. If it is the engagement of the verbal system in particular that disrupts learning, one would expect that speech production would lead to greater disruptions in word learning compared to other non-verbal tasks, such as performing a spatial task. This is currently being addressed in our research.

Our results suggest that a re-evaluation is needed for how production is conceived in language development, taking into account stimuli complexity (Kaushanskaya & Yoo, 2011; Zamuner et al., submitted), task complexity, and the learners' development. The notion that the effects of production may vary has been previously noted. Vihman et al. (2014) and Vihman (2017) discuss how the articulatory filter (previously produced sounds are more salient in the learner's input) may show differential effects depending on attentional resources, processing demands of the task, and the developmental stage of the learner. Young children are more likely to learn words with highly familiar meanings and/or sounds (also familiar in that the sounds are produced by the child) because this requires fewer processing demands, and allows for more resources for mapping a word form and its referent (Vihman et al., 2014, p. 127). When actively recalling a word, the processing demands are thought to be high, and advantages will be seen for items within the children's existing production repertoire. However, once the learner has mastered production patterns, in less demanding tasks such as passive listening, listeners attend more to patterns not in their production repertoire. As we hypothesize that producing a novel word form during word-reference mapping has high task demands, this suggests that it may be more efficient for learners to initially passively listen to language before engaging the production system, freeing up resources for the creation of phonological, lexical, and semantic representations. This is expected to be even more relevant when learners are acquiring words with sound patterns that have yet to be mastered in production, because a mismatch between a child's output and target would leave even fewer processing resources and disrupt the mapping even more.

At the same time, speech production can be informative in other ways, because it can provide articulatory information and kinetic feedback to the learner (McAllister Byun et al., 2016). There is also variability in the production skills of young children, and not all children may benefit from producing or hearing words in the same way. This highlights the fact that the effect of hearing and producing on children's word recognition may vary depending on the level of representation probed in the task. Recognition in our study was tested using the same auditory tokens presented during training. Other directions for future research would be to test children's abilities to detect slight mismatches in targets (e.g., *kel* produced as *zel*), or to test children's ability to generalize to new speakers. Another area for future research would be to manipulate the lexical status of the new words. In Heisler et al. (2010), children were taught word forms either with or without a visual referent. In the former condition (word forms with a visual referent), children produced more stable articulations. This suggests that articulatory representations are not independent from lexical representations. We predict that if children were taught novel words without visual referents, using the same paradigm of produced and heard during training, that in this type of simpler task one might find the classic production effect. It would also be interesting to examine whether manipulations of lexical status, with and without production during training, would result in differences in short-term versus long-term learning. This is based on work by Hodges et al. (2016) who found that the accuracy of 2.5- to 3-year-old children's novel word imitation was related to children's performance in a cued recall task, following 1-minute and 5-minutes after training. However, children's imitation accuracy during training was not predictive of their long-term performance on the cued recall task, which was tested 1 to 7 days past training. This suggests that imitative productions might have an impact on children's short-term learning,

but that other factors play a role in integrating lexical representations into long-term memory. Related to this, future research may explore the difference between imitation versus retrieval, where retrieval has been shown to have a strong effect on second language vocabulary learning with adults (Kang, Gollan, & Pashler, 2013).

One last consideration is children's language background and language skills. We acknowledge that it would have been more accurate and potentially informative for us to have measured children's language background and skills using a standardized test, rather than using parental reports. Our reliance on parental reports was based on practical reasons of time limitations inherent in testing at a museum-based lab. Many of the children in our study were reported as having exposure to another language. The majority of these children were from monolingual English-speaking households, who had started receiving 30 minutes of French a day in kindergarten at 4 years of age. This is the typical profile of children who are starting French immersion in Ottawa, Canada, where the participants were tested. While we consider the participants in our experiments as functionally monolingual, a child's language background could have influence on the direction of the production effect. For example, a child who is an L2 speaker of the test language could show the reverse production effect if the test stimuli do not map to the child's L1 phonology (as seen with adults in Kaushanskaya & Yoo, 2011). Individual children's language skills might also influence their performance, which could be measured in different ways, such as using standardized measures of articulation skills (e.g., Goldman-Fristoe's Test of Articulation), receptive vocabulary size (e.g., Peabody Picture Vocabulary Test) or expressive vocabulary size (e.g., Expressive One-Word Picture Vocabulary Test). Lastly, work is needed to look at the effect of production across development as we predict that this may vary—showing different effects for toddlers compared to school-aged children. Future research could address this by extending the current research to children as young as 2 years of age. The success of the block design in Experiment 2 provides a promising starting point for future research that could span different age groups.

Across two experiments, we found that 4.5- to 6-year-old children showed differences in their recognition and recall of novel words that were heard compared to produced during a brief training phase. To acquire spoken language and communicate, learners must move beyond their initial abilities, develop vocalic motor skills, and unify perception and production (Davis, MacNeilage, & Matyear, 2002; Pierrehumbert, 2003; Vihman, DePaolis, & Keren-Portnoy, 2009). It is an open question as to how the emerging speech production system interacts with other factors, such as linguistic complexity, the articulation skills of the learner, the complexity of the task, among others. These types of future studies are needed to fully understand the role of production in language development.

ACKNOWLEDGEMENTS

This research was supported by an SSHRC grant awarded to TSZ and an FQRSC Doctoral Fellowship awarded to EML. We thank the Canada Aviation and Space Museum where a portion of the data was collected.

ENDNOTES

[1] One possibility raised by a reviewer is that the difference of effect between Experiment 1 and Icht and Mama (2015) might stem from the gender breakdown of our participants. In our Experiment 1 there were 12 males and 4 females. In comparison, the gender distribution in Icht and Mama was more balanced (Experiment 1: 18 males, 12 females; Experiment 2: 13 males, 17 females). Previous research has found that language skills are slower to develop in boys compared to girls (Eriksson et al., 2012). Therefore, it is possible that the reverse production effect in our study stemmed from the disproportionate number of boys that we had, who may have had less developed language skills. To test for this possibility, we conducted the same GCA analysis, but with the additional factors of gender and age using a median split (younger, older). We also performed the same analysis with age as a continuous variable and the results were the same. There were no significant effects of gender, age, gender and training, and age and training. This means that, overall, males or females in our study did not look more to heard targets than produced targets, and that younger or older participants did not look more to heard targets than produced targets. However, there were two interactions: An interaction of training and gender on the linear term (Estimate = −5.62, SE = 2.00, p = .004), and an interaction of training and age on the linear term (Estimate = −9.12, SE = 2.94, p = .002). This reflects that the slope of females' looking on heard trials was steeper than the slope of males, and that younger children's slope for looking on heard trials was steeper than older children's looking slope on heard trials. These effects need to be interpreted with caution because gender and age were not controlled variables, and they may not be true effects due to low numbers.

[2] As in Experiment 1, the distribution of males/females in Experiment 2 was not balanced (12 males, 4 females), so we conducted the same GCA analysis, but with the additional factors of gender and age, with age treated a median split. (The results were the same when age was treated as a continuous variable.) As in Experiment 1, for Experiment 2 there were no significant effects of gender, age, gender and training, and age and training. This means that males or females did not look more to heard targets than produced targets overall and that younger or older participants did not look more to heard targets than produced targets. However, there were three significant interactions. The first interaction was a significant effect of gender and training on the linear term (Estimate = 4.99, SE = 1.44, p < .001), indicating that males' heard looking curve was flatter than that of females. The second interaction was a significant effect of age and training on the linear term (Estimate = 7.44, SE = 2.41, p = .002), indicating that older children's heard slope was steeper than that of younger children. The third interaction involved a significant interaction between training, gender, age, and the linear term (Estimate = −10.90, SE = 2.88, p < .001). This interaction effect reflects that females' slopes for heard trials were steeper than males' slopes, and that older children's slopes were steeper than younger children's. Note that these additional analyses need to be interpreted cautiously and to know whether these effects are true effects, future research would need to experimentally control gender and age.

REFERENCES

Abbs, B., Gupta, P., & Khetarpal, N. (2008). Is overt repetition critical to expressive word learning? The role of overt repetition in word learning with and without semantics. *Applied Psycholinguistics*, *29*, 627–667.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. *VIII* (pp. 47–90). New York: Academic Press.

Baddeley, A.D., Gathercole, S.E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.

Baese-Berk, M.M., & Samuel, A.G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36.

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. Amsterdam: John Benjamins.

Barr, D.J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7. http://CRAN.R-project.org/package=lme4.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, USA*, *109*, 3253–3258.

Bruderer, A.G., Danielson, D.K., Kandhadai, P., & Werker, J.F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences, USA*, *112*, 13531–13536.

Buckler, H., & Fikkert, P. (2016). Dutch and German 3-year-olds' representations of voicing alternations. *Language and Speech*, *59*, 236–265.

Carlson, M.T., Sonderegger, M., & Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory and Language*, *75*, 159–180.

Cho, K.W., & Feldman, L.B. (2013). Production and accent affect memory. *The Mental Lexicon*, *8*, 295–319.

Cho, K.W., & Feldman, L.B. (2016). When repeating aloud enhances episodic memory for spoken words: Interactions between production- and perception-derived variability. *Journal of Cognitive Psychology*, *28*, 673–683.

Clark, E.V. (2007). Young children's uptake of new words in conversation. *Language in Society*, *36*, 157–182.

Coady, J.A., & Aslin, R.N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*, *89*, 183–213.

Conway, M.A., & Gathercole, S.E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*, 341–361.

Core, C. (2012). Assessing phonological knowledge. In E. Hoff (Ed.), *Research methods in child language: A practical guide* (pp. 77–99). Oxford: Wiley-Blackwell.

Curtin, S., Byers-Heinlein, K., & Werker, J.F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*, 492–504.

Curtin, S., & Zamuner, T.S. (2014). Understanding the developing sound system: Interactions between sounds and words. *WIREs Cognitive Science*, *5*, 589–602.

Dahlen, K., & Caldwell-Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *Modern Languages Journal*, *97*, 902–916.

Davis, B.L., MacNeilage, P.F., & Matyear, C. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica*, *59*, 75–107.

Dell, G.S., Schwartz, M.F., Nozari, N., Faseyitan, O., & Coslett, H.B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *3*, 380–396.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, *49*, 137–173.

DePaolis, R.A., Vihman, M.M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, *34*, 590–601.

DePaolis, R.A., Vihman, M.M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, *36*, 642–649.

Eriksson, M., Marschik, P.B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., … Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, *30*, 326–343.

Ettlinger, M., Lanter, J., & Van Pay, C.K. (2014). Learning to remember by learning to speak. *Developmental Psychology*, *50*, 431–438.

Fawcett, J.M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, *142*, 1–5.

Fennell, C.T. (2012). Object familiarity enhances infants' use of phonetic detail in novel words. *Infancy*, *17*, 339–353.

Fennell, C.T., & Werker, J.F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, *46*, 245–264.

Ferguson, C.A., & Farwell, C.B. (1975). Words and sounds in early language acquisition. *Language*, *51*, 419–439.

Gathercole, S.E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*, 513–543.

Gathercole, S.E., & Conway, M.A. (1988). Exploring long-term modality effects: Vocalization leads to better retention. *Memory and Cognition*, *16*, 110–119.

Gathercole, S.E., Pickering, S.J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*, 177–190.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Gupta, P., & Tisdale, J. (2009). Word learning, phonological short-term memory, phonotactic probability and long-term memory: Towards an integrated framework. *Philosophical Transactions of the Royal Society B*, *364*, 3755–3771.

Heisler, L., Goffman, L., & Younger, B. (2010). Lexical and articulatory interactions in children's language production. *Developmental Science*, *13*, 722–730.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*, 135–145.

Hodges, R., Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2016). The role of elicited verbal imitation in toddlers' word learning. *Journal of Child Language*, *43*, 457–471.

Hopkins, R.H., & Edwards, R.E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behaviour*, *11*, 534–537.

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

Kan, P.F., & Sadagopan, N. (2015). Speech practice effects on bilingual children's fast mapping performance. *Seminars in Speech and Language*, *36*, 109–119.

Kan, P.F., Sadagopan, N., Janich, L., & Andrade, M. (2014). Effects of speech practice on fast mapping in monolingual and bilingual speakers. *Journal of Speech, Language, and Hearing Research*, *57*, 929–941.

Kang, S.H., Gollan, T.H., & Pashler, H. (2013). Don't just repeat after me: Retrieval practice is better than imitation for foreign vocabulary learning. *Psychonomic Bulletin and Review*, *20*, 1259–1265.

Kaushanskaya, M., & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, *26*, 121–148.

Keren-Portnoy, T., Vihman, M.M., DePaolis, R.A., Whitaker, C.J., & Williams, N.M. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language, and Hearing Research*, *53*, 1280–1293.

Leonard, L.B., Schwartz, R.G., Morris, B., & Chapman, K. (1981). Factors influencing early lexical acquisition: Lexical orientation and phonological composition. *Child Development*, *52*, 882–887.

McAllister Byun, T., Inkelas, S., & Rose, Y. (2016). The A-map model. *Language*, *92*, 141–178.

McAllister Byun, T., & Tessier, A.-M. (2016). Motor influences on grammar in an emergentist model of phonology. *Language and Linguistic Compass*, *10*, 431–452.

MacLeod, C.M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin and Review*, *18*, 1197–1202.

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.

Maekawa, J., & Storkel, H.L. (2006). Individual differences in the influence of phonological characteristics on expressive vocabulary development by young children. *Journal of Child Language*, *33*, 439–459.

Majorano, M., Vihman, M.M., & DePaolis, R.A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, *10*, 179–204.

Mama, Y., & Icht, M. (2016). Auditioning the distinctiveness account: Expanding the production effect to the auditory modality reveals the superiority of writing over vocalising. *Memory*, *24*, 98–113.

Mathias, B., Palmer, C., Perrin, F., & Tillmann, B. (2015). Sensorimotor learning enhances expectations during auditory perception. *Cerebral Cortex*, *25*, 2238–2254.

Matin, E., Shao, K.C., & Boff, K.R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception and Psychophysics*, *53*, 372–380.

Mirman, D., Dixon, J.A., & Magnuson, J.S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*, 475–494.

Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, *3*, 41.

Munson, B., Edwards, J., & Beckman, M.E. (2012). Phonological representations in language acquisition: Climbing the ladder of abstraction. In A. Cohn, S. Fougeron & M. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 288–309). Oxford: Oxford University Press.

Ngon, C., & Peperkamp, S. (2016). What infants know about the unsaid: Phonological categorization in the absence of auditory input. *Cognition*, *152*, 53–60.

Ohala, D.K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders*, *32*, 397–422.

Ota, M., & Green, S.J. (2013). Input frequency and lexical variability in phonological development: A survival analysis of word-initial cluster production. *Journal of Child Language*, *40*, 539–566.

Ozubko, J.D., & MacLeod, C.M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1543–1547.

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.

Page, M.P.A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *364*, 3737–3753.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*, 115–154.

Salverda, A.P., Kleinschmidt, D., & Tanenhaus, M.K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*, 145–163.

Schwartz, R.G., & Leonard, L.B. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, *9*, 319–336.

Sosa, A.V., & Bybee, J.L. (2008). A cognitive approach to clinical phonology. In M.J. Ball, M.R. Perkins, N. Müller, & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 480–490). Oxford: Blackwell.

Sosa, A.V., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research*, *55*, 596–608.

Stoel-Gammon, C. (1998). Sounds and words in early language acquisition: The relationship between lexical and phonological development. In R. Paul (Ed.), *Exploring the speech–language connection* (pp. 25–52). Baltimore, MD: Paul H. Brookes.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*, 1–34.

Storkel, H.L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, *25*, 201–221.

Storkel, H.L. (2006). Do children still pick and choose? The relationship between phonological knowledge and lexical acquisition beyond 50 words. *Clinical Linguistics and Phonetics*, *20*, 523–529.

Storkel, H.L. (2013). A corpus of consonant-vowel-consonant real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods*, *45*, 1159–1167.

Storkel, H.L., Bontempo, D.E., Aschenbrenner, A.J., Maekawa, J., & Lee, S.Y. (2013). The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Speech, Language, and Hearing Research*, *56*, 1689–1700.

Storkel, H.L., & Hoover, J.R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, *42*, 497–506.

Swingley, D., & Aslin, R.N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147–166.

Vihman, M.M. (1993). Variable paths to early word production. *Journal of Phonetics*, *21*, 61–82.

Vihman, M.M. (2017). Learning words and learning sounds: Advances in early lexical and phonological development. *British Journal of Psychology*, *108*, 1–27.

Vihman, M.M., DePaolis, R.A., & Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning*, *64*, 121–140.

Vihman, M.M., Macken, M.A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language*, *61*, 397–445.

Vihman, M.M., DePaolis, R.A., & Keren-Portnoy, T. (2009). Babbling and words: A dynamic systems perspective on phonological development. In E.L. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 163–182). Cambridge: Cambridge University Press.

Werker, J.F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197–234.

Yeung, H.H., & Werker, J.F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, *245*, 603–612.

Zamuner, T.S., Fikkert, P., & Gick, B. (2007). Production facilitates lexical acquisition in young children. *Journal of the Acoustical Society of America*, *122*, 3031.

Zamuner, T.S., Morin-Lessard, E., Strahm, S., & Page, M.P. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language*, *89*, 55–67.

Zamuner, T.S., Strahm, S., Morin-Lessard, E., & Page, M.P.A. (submitted). The effect of production on the recognition and recall of novel words depends on phonotactic probability. Manuscript under review.

Zamuner, T.S., & Thiessen, A. (in press). A phonological, lexical, and phonetic analysis of the new words that young children imitate. *Canadian Journal of Linguistics*.

Zamuner, T.S., Yeung, H.H., & Ducos, M. (2017). The many facets of speech production and its complex effects on phonological processing. *British Journal of Psychology*, *108*, 37–39.

How to cite this article: Zamuner TS, Strahm S, Morin-Lessard E, Page MPA. Reverse production effect: children recognize novel words better when they are heard rather than produced. *Dev Sci*. 2018;21:e12636. https://doi.org/10.1111/desc.12636