JSLHR

Research Article

Frequencies in Perception and Production Differentially Affect Child Speech

Peter T. Richtsmeier[a] and Amanda K. Good[a]

Purpose: Frequent sounds and frequent words are both acquired at an earlier age and are produced by children more accurately. Recent research suggests that frequency is not always a facilitative concept, however. Interactions between input frequency in perception and practice frequency in production may limit or inhibit growth. In this study, we consider how a range of input frequencies affect production accuracy and referent identification.

Method: Thirty-three typically developing 3- and 4-year-olds participated in a novel word-learning task. In the initial test block, participants heard nonwords 1, 3, 6, or 10 times—produced either by a single talker or by multiple talkers—and then produced them immediately. In a posttest, participants heard all nonwords just once and then produced them. Referent identification was probed in between the test and posttest.

Results: Production accuracy was most clearly facilitated by an input frequency of 3 during the test block. Input frequency interacted with production practice, and the facilitative effect of input frequency did not carry over to the posttest. Talker variability did not affect accuracy, regardless of input frequency. The referent identification results did not favor talker variability or a particular input frequency value, but participants were able to learn the words at better than chance levels.

Conclusions: The results confirm that the input can be facilitative, but input frequency and production practice interact in ways that limit input-based learning, and more input is not always better. Future research on this interaction may allow clinicians to optimize various types of frequency commonly used during therapy.

Frequency is an important concept across all aspects of language acquisition (Ambridge, Kidd, Rowland, & Theakston, 2015). At the same time, given its sundry applications, "frequency" can mean so many different things that it risks becoming uninformative. In this study, we begin with an overview of frequency effects in expressive language development. Studies of speech accuracy and word learning reveal consistent frequency effects. We then focus more narrowly on two types of frequency that are fundamental to speech and language development. The first type is input frequency or a perceptual measure of how often children hear a target form. The second type is production practice or an expressive measure of how often children produce a target form. Both types of frequency are beneficial, but emerging evidence suggests that some combinations of input frequency and production practice can inhibit production accuracy and word learning. We then report an experiment with 3- and 4-year-old children that examines both input frequency and production practice in novel word production and form-referent learning.

Frequency in Expressive Language Development

Frequent sounds are at an advantage during speech acquisition. For example, Stoel-Gammon (1998) observes that children's first words tend to be composed of frequent syllable shapes and phonemes, and they conform to frequent stress patterns. As children continue to develop into the preschool years, they are more accurate when producing words and many of the sound sequences that are frequent in their language (Edwards & Beckman, 2008; Masdottir & Stokes, 2016). Similarly, the first nouns and verbs that children acquire tend to be high frequency (Goodman, Dale, & Li, 2008; Tatsumi, Ambridge, & Pine, 2017).

Frequency also relates to the errors that children make. Stemberger and Bernhardt (1999) describe how frequent, early acquired sounds may act as default sounds, replacing less frequent sounds in child speech patterns such as fronting ("key" pronounced [ti]). The alveolar stops /t, d/ are more frequent cross-linguistically than /k, g/, and they are

[a]Department of Communication Sciences and Disorders, Oklahoma State University, Stillwater

Amanda K. Good is now a Speech-Language Pathologist at Canyon Breeze Elementary School, Avondale, AZ.

Correspondence to Peter T. Richtsmeier: prichtsmeier@yahoo.com

Editor-in-Chief: Julie Liss
Editor: Bharath Chandrasekaran

Received October 17, 2017
Revision received April 9, 2018
Accepted July 16, 2018
https://doi.org/10.1044/2018_JSLHR-S-17-0391

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

also more frequent in English. Thus, frequency helps explain why children substitute the more frequent alveolars for velars, although factors such as ease of articulation are also relevant (for a discussion, see Stokes & Surendran, 2005). In the domain of word learning, frequent words tend to stand in for infrequent words in semantic overgeneralization errors (Ambridge, Pine, Rowland, Chang, & Bidgood, 2013; see Gershkoff-Stowe, Connell, & Smith, 2006, for a discussion of both frequency and repetition priming).

Frequency is a prominent explanatory factor in expressive language development, yet it is often more beneficial to distinguish among different types of frequency. In the following sections, we emphasize a distinction between input frequency and production practice, the former being how often children hear a target form and, the latter, how often they produce it. As we will see, the literature suggests that both input frequency and production practice play an important role in understanding how and when increased frequency is facilitative.

Input Frequency

We begin with a review of studies that emphasize input frequency. Using a corpus of early child speech, Ota and Green (2013) asked whether the frequency that children heard word-initial consonant clusters would predict how quickly children produced those clusters accurately. The corpus included productions of both two- and three-consonant clusters from three children; recordings began around each child's first birthday and stopped around their fourth birthday. Input frequency was a significant predictor of accuracy even when controlling for related factors such as the number of attempts the children made at producing the clusters. Zamuner, Gerken, and Hammond (2004) report that 2-year-olds produce coda consonants more accurately when they occur in more frequent phonotactic environments. For example, children produced /d/ more accurately in the relatively probable nonword /gɛd/ than in the less probable /tʃʌd/.

Regarding word learning, Goodman et al. (2008) compared corpora of child word acquisition and word frequency in child-directed speech. They estimated the age of acquisition for several hundred words with data from the MacArthur–Bates Communicative Development Inventories (Fenson et al., 2007); the input frequency of those same words was estimated using the Child Language Data Exchange System database (where *input frequency* was defined as presence in child-directed speech). Age of acquisition was clearly mediated by a word's semantic/syntactic category, but within each category, input frequency was a robust predictor.

In a series of influential studies, Storkel (2001, 2003) examined learning of novel nouns and verbs in young children. The novel words were controlled for phonotactic probabilities, with some words being composed of frequent sound sequences and others composed of infrequent sequences. Children tended to learn both nouns and verbs better if they were composed of frequent sound sequences. Importantly for this study, children were also sensitive to the number of exposures of the novel words and typically exhibited greater learning with additional exposures, indicating an important role for input frequency.

Tatsumi et al. (2017) looked at the predictive capacity of input frequency in the acquisition of inflected verbs in Japanese. The authors note that more frequent verb forms tend to have simpler inflections, but the exact correlation changes from verb to verb. Preschool-aged children acquiring Japanese were given pictures to describe, a target verb, and a prompting phrase. The authors found that children's choice of a simple or complex inflection was predicted by the relative frequency of each form in child-directed speech. Thus, input frequency helps explain why children will sometimes produce more complex forms of a verb.

Although input frequency by itself is often a valid predictor of speech sound development, certain aspects of a frequent input appear to facilitate learning. For instance, Plante, Bahl, Vance, and Gerken (2011) demonstrated that input variability—in the form of exposure to multiple talkers—is an effective component of input-based learning of novel words. In this study, typically developing children and children with specific language impairment were familiarized with nonwords. Children heard half of the words in a high variability condition in which the nonwords were produced by multiple talkers; they heard the other nonwords produced by a single talker in a low variability condition. The high variability condition increased production accuracy and reduced response times in both groups of children. The authors conclude that variability, such as across talkers, may be as important as the raw input frequency of the learning target.

Other experimental work finds that the right mix of variables in the input can allow children to generalize to new words. One example of an input variable is variability across related words, or word-type frequency, which is thought to support linguistic category learning (Pierrehumbert, 2003). For example, Richtsmeier, Gerken, and Ohala (2011) examined the relative contributions of talker variability and word-type frequency to generalization. Children were significantly more accurate to produce target consonant sequences when they were first exposed to those sequences through a combination of multiple talkers and multiple word types. The authors conclude that "high frequency" likely represents a combination of variables, including factors such as multiple talkers and word types.

Input frequency is also a critical component to several approaches to speech and language therapy. Regarding speech sounds, the Speech Assessment and Interactive Learning System (SAILS) developed by Rvachew and colleagues (for example, Rvachew & Brosseau-Lapré, 2015) is used to help children distinguish between the errors that they make on a target sound and correct productions of that sound. Additionally, both the traditional method of articulation therapy (Van Riper & Erickson, 1996) and the cycles approach (Hodson & Paden, 1991) call for clinicians to employ perceptual training as a part of therapeutic sessions.

Regarding word learning, Storkel et al. (2017) conducted a treatment study of interactive book reading in which the goal was to identify the adequate teaching intensity for children with specific language impairment to be able to learn new words. Based on groups of children receiving 12, 24, 36, and 48 exposures, 36 exposures were adequate for learning to be observed. In sum, both basic and clinically oriented research have determined that input frequency can have a facilitative effect on speech production and word learning.

Production Practice

Production practice appears to play a role in children's earliest word productions. Vihman and others have shown that frequently babbled sounds and syllables become the sounds and syllables of first words (Robb & Bleile, 1994; Vihman, 1992). The implication is that babbling is a kind of production practice, allowing infants to develop the motor skills necessary to articulate sounds intentionally in their first words.

Two complementary lines of research indicate that production practice supports the growth of children's speech sound inventories. In one line of research, Stoel-Gammon and Dale (1988) examined a group of precocious talkers whose expressive vocabularies at 18 months were as much as 10 times the size of the average 18-month-old's vocabulary. By acquiring so many words, these children also appeared to have expanded their phonetic inventories. In the other line of research, Rescorla and Ratner (1996) worked with late talkers—who produce very few words—and revealed a correlation between low production practice and limited phonetic inventories.

In her analysis of the link between the early lexicon and production accuracy, Stoel-Gammon (1998) proposes a central role for production practice. Using the MacArthur–Bates Communicative Development Inventories, she examined the correlation between (a) the frequency of sounds on the inventory's expressive vocabulary list and (b) production accuracy for that sound in a normative study (Templin, 1957). There was a robust correlation of .71 between word-initial sound frequency and production accuracy and a modest correlation of .53 between word-final sound frequency and production accuracy. One plausible explanation for these correlations is that production practice with an emerging lexicon drives the growth in production accuracy.

Turning to word learning, children's expressive lexicons grow rapidly, and production practice appears to support that growth. Gershkoff-Stowe (2002) conducted two experiments on word retrieval during the second and third years of life. In the first experiment, 1-year-olds were exposed to two sets of words, one that they practiced frequently and the other that they practiced a few times. Children typically made fewer errors when naming the frequently practiced words. In the second experiment, 2-year-olds were exposed to two picture books, one with more commonly produced animals and one with less commonly produced animals. Congruent with the first experiment, children tended to make more errors in the book containing less commonly produced animals. Although Gershkoff-Stowe did not hold input frequency constant in either experiment, her results suggest that production practice supports accurate word retrieval in children's expanding expressive vocabularies.

Production practice supports lexical development in other ways, as well. For example, Zamuner and Thiessen (in press) observed in a corpus of child speech that children tended to acquire new words with existing sounds from their expressive speech sound inventories. Production practice may also make words with practiced sounds more salient (Vihman, DePaolis, & Keren-Portnoy, 2014) and improve fast mapping of new words in bilingual preschoolers (Kan & Sadagopan, 2015).

The clinical literature indicates that production practice is important for understanding how children with speech delays can improve production accuracy. Maas et al. (2008) discuss how production practice should be carefully considered to help a client improve speech accuracy (but see Maas, Butalla, & Farinella, 2012; Maas & Farinella, 2012, for a discussion of individual differences). Edeal and Gildersleeve-Neumann (2011) manipulated production practice in a treatment study of two children with childhood apraxia of speech. Both children made significantly more progress for sounds in the high production practice treatment condition. In sum, production practice appears to have a direct influence on furthering speech and language development. As with input frequency, studies that isolate production practice generally find it to be facilitative of a child's development.

Interactions Between Perceptual and Production Frequencies

The previous sections considered input frequency and production practice in isolation, but natural language acquisition necessarily involves both in combination. In fact, a growing literature suggests that speech sound development is subject to interactions between what is heard and what is said.

A longstanding proposal for the interaction of perception and production across development involves learning by imitation. Proposals by Guenther (1995), Hewlett (1990), and Kuhl (2000), for example, argue that children learn to produce speech sounds by imitating categories learned via perception. Imitation may also benefit from a perceptual feedback loop. In the feedback loop, children increase production accuracy based on a perceptual evaluation of their own productions (see Hickok, 2014, for an instantiation of this feedback loop). In all these proposals, perceptual experience acts as a starting point and as a form of corrective feedback, whereas production practice acts as the driving force, compelling the production mechanism forward in development. In sum, research on imitation indicates that perception and production function synergistically to support increased production accuracy over time.

Emerging research suggests that perception and production do not always benefit one another, however. For example, Kaushanskaya and Yoo (2011) and Baese-Berk and Samuel (2016) both observed processing deficits for nonnative sounds when adult participants produced those sounds compared with when they just listened to them. In a study with 4- to 6-year-old children, Richtsmeier and Goffman (2017) found that production practice limited perceptual phonological learning in children. Participants were familiarized with word-medial consonant sequences such as /fp/ in the context of novel CVCCVC (C = consonant; V = vowel) words. Children were either exposed to multiple words produced by multiple talkers (e.g., /fp/ in /nɪfpən/, /ʃeɪfpəs/, and /kofpət/) in a high variability condition or to a single word produced by a single talker in a low variability condition. Learning was assessed immediately and 1 week later via production of related test words (e.g., /mæfpəm/). At each time point, nine productions of each test word were collected. Children only showed an advantage for the high variability familiarization in their first few productions in the immediate test. The authors interpret the limited evidence for perceptual learning to reflect a normalizing effect of production practice, reducing the benefits of the high variability familiarization with each additional production.

Another proposal in which perception and production are not completely synergistic is the A-map proposed by McAllister Byun, Inkelas, and Rose (2016). The A-map suggests that production and perception exert distinct influences on the developing child. In terms of production, immature articulatory and motor systems prevent children from producing accurate speech. At the same time, consistent productions are favored, even when the production target is incorrect. In contrast, perceptual learning from the input establishes phonological categories that match adult targets and that children's productions should eventually match. Thus, the articulators and motor system initially prevent children from achieving productions corresponding to their input-derived phonological categories. Once the articulatory and motor systems have matured, however, children can more easily adapt their productions to match the categories they have learned via perception.

Production practice may also have costs for word learning. In a study with 4.5- to 6-year-old children, Zamuner, Strahm, Morin-Lessard, and Page (2017) used an eye-tracking paradigm to examine word learning with and without production practice. They familiarized children with novel CVC words (e.g., /kɛl/) paired with make-believe animals. For half of the words, familiarization required hearing the make-believe animal name and then repeating it. For the other half, familiarization required children to hear the names twice, thus equating input frequency across both types of familiarization. In a forced-choice looking task, children looked longer at the animals whose names they heard relative to the animals whose names they produced. Children were also more likely to name the animals they heard in an open-ended recall task. Zamuner et al. (2017) argue that there are processing costs associated with production when young children are engaged in word learning.

Independently, perceptual and production frequencies appear to benefit speech production and word learning. However, recent research suggests that perception and production do not always have simple, additive effects. Research going forward should help us understand how perception and production combine to either facilitate or inhibit speech development.

Current Study

In the introduction, we emphasized speech modalities and the fact that the acts of listening and speaking represent separable types of frequencies. Furthermore, the combination of input frequency and production practice sometimes inhibits speech accuracy and word learning. Given this complexity, the present learning study was conducted to answer persistent questions about how listening and speaking interact.

Is More Always Better?

As reviewed above, the literature on input frequency effects is generally a collection of facilitative effects, with greater frequencies facilitating more perceptual learning. A "more is always better" interpretation of input frequency effects is also consistent with the literature on talker variability, in which exposure to multiple talkers is facilitative. Nevertheless, explicit, experimental comparisons of multiple levels of frequency are lacking.

From a clinical perspective, it is also important to understand the effect of frequency within a single sitting. Although broad developmental trends suggest that the most frequent sounds and sound sequences are acquired faster (Edwards & Beckman, 2008), it does not necessarily follow that the most frequently targeted sound within a single therapy session will improve the most. The distinction between broad frequency effects and their implementation within a session is discussed in depth by Warren, Fey, and Yoder (2007). Based on their terminology, this study primarily addresses *dose*, which Warren et al. define as "the number of properly implemented teaching episodes during a single intervention session" (p. 71).

In this study, we exposed typically developing children to nonwords with four different input frequencies: one, three, six, and 10. If more is always better, we would expect a linear increase in speech accuracy corresponding to each increase in input frequency: 1 > 3 > 6 > 10. Similarly, we might expect more robust word learning with greater input frequencies. A note of caution is appropriate here, however, as raw frequency appears to be less influential for word learning than other factors such as semantic depth (Capone & McGregor, 2005; Gladfelter & Goffman, 2018) that we do not address.

In previous studies of perceptual learning, aspects of a frequent input such as talker variability facilitated learning (Plante et al., 2011; Richtsmeier, Gerken, Goffman, & Hogan, 2009). However, those studies used a between-subjects design in which separate groups of participants participated in conditions with and without talker variability.

Here, we used a within-subjects design to examine whether talker variability supports perceptual learning beyond raw frequency.

How Do the Effects of Input Frequency Play Out Across Repeated Productions?

In natural language, perceptual learning must intermingle with speaking opportunities. Therefore, it is important to understand how input frequency, which is often observed to be facilitative, interacts with production practice. Previous experimental studies suggest that production practice may impede perceptual learning (Zamuner et al., 2017) or limit it (Richtsmeier & Goffman, 2017). However, those studies used a design in which training and testing were separately blocked, with the learning only occurring at the beginning of the experiment. Those designs may therefore be subject to unintended influences from working memory. In this study, participants were exposed to the target input frequencies immediately before each production. This design limits the need for a learned pattern to be maintained in memory, and changes to speech accuracy can be more readily attributed to either input frequency, production practice, or both.

Longer term learning was also considered in the design. Following a test session in which input frequencies were manipulated, children completed several other tasks, and then, they produced the target nonwords in a posttest. The purpose of the posttest was to determine whether perceptual learning would be maintained over time, but it also provided an opportunity to examine additional opportunities for production practice.

Method

Participants

Forty-one children between the ages of 3;0 and 4;8 (years;months) were recruited for the experiment (M_{age} = 3;10). Participants were recruited through daycares and preschools from the surrounding area. Advertisements in local newspapers were also used. In all cases, parents voluntarily contacted the researchers. All participants were monolingual native English speakers and met criteria for typical development, including passing a standardized test of articulation, a hearing screening, and other criteria obtained through a parent questionnaire. Children completed the Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2; Goldman & Fristoe, 2000), and all received a standard score of 85 or above. The average score was 113, indicating that most children had above-average articulation skills. All children passed a hearing screening by indicating that they heard pure tones in each ear of 1000, 2000, and 4000 Hz at 25 dB. Using a questionnaire completed by parents, the participants' developmental history was reviewed. All but two children were reported as having typical development. Two had a history of speech and language therapy and were excluded from the analyses. Children who participated also completed an auditory discrimination task—described below—and a non–word repetition task based on Dollaghan and Campbell (1998). Six children were unable or unwilling to complete the experiment and are not included in the analyses. The remaining 33 children, 18 male and 15 female, were included. The results of the GFTA-2, the auditory discrimination task, and the non–word repetition task are summarized in Table 1.

Materials

Materials included eight CVCCVC nonwords: /pɛmtəs/, /niʃkət/, /mæfpəg/, /fugdən/, /sabləf/, /tʌvʧəp/, /bozjəm/, and /gɪsnək/ (similar nonwords have been used elsewhere, e.g., Munson, 2001; Richtsmeier et al., 2009). Sixteen sounds were included in the word-medial consonant sequences, and no sound was repeated in more than one word-medial sequence. For example, in /pɛmtəs/, the word-medial /m/ and /t/ do not appear medially in the other seven nonwords. Initial and final sounds of the nonwords were selected subsequently to meet two goals. First, the initial stressed syllable should not form a real word. Second, the full nonword should have no neighbors based on a search of the The Irvine Phonotactic Online Dictionary database (Vaden, Halpin, & Hickok, 2009). Although the learning targets were whole words, medial consonant sequences were carefully considered because children are likely to make errors when producing them (McLeod, 2016; Smit, 1993). Therefore, medial consonant sequences create an opportunity for production accuracy to change, allowing us to infer learning. One aspect of the consonant sequences that was left uncontrolled was phonotactic probability, which was allowed to vary. Table 2 provides a summary of the phonotactic probabilities of all eight sequences. Typically, children are more accurate when producing high-probability sequences (Munson, 2001; Richtsmeier et al., 2009). It was important to counterbalance the eight words to manage this variability, and the assignment of words to experimental conditions is described at the end of this section.

The nonwords were placed in four input frequency conditions, each differing in the number of familiarization exposures immediately before each production. Two of the nonwords were heard one time, two were heard three times, two were heard six times, and two were heard 10 times. The existing literature suggests that greater frequency is generally better, but it is unclear whether that generalization holds for shorter time spans. These four levels of input frequency allow us to test this hypothesis across a time span of about 1 hr.

The nonwords were also divided into two talker variability conditions, in which either a single talker or multiple talkers presented each nonword. Because input frequency varied from one to 10, the total number of talkers also varied as a consequence of input frequency. One example of the different total number of talkers appears in Figure 1. In that figure, /gɪsnək/ was presented six times by six different talkers, whereas /mæfpəg/ was presented 10 times by 10 different talkers. Now comparing /mæfpəg/ to /pɛmtəs/, both had an input frequency of 10 but differed in terms of talker variability. By default, the two nonwords in the input frequency level of 1 were heard by a single talker.

Table 1. Mean scores, standard deviations, and confidence intervals for the descriptive measures collected from participants.

Descriptive measure	*n*	*M* (*SD*)	95% CI
Age in months	33	45.67 (6.06)	[43.52, 47.81]
Goldman-Fristoe Test of Articulation–2 standard scores	33	112.72 (7.62)	[110.02, 115.43]
Auditory discrimination task	24	75% (17%)	[68%, 82%]
Non–word repetition	24	75% (13%)	[69%, 80%]

Note. Nine children were unable to complete the auditory discrimination task. Three children were unable to complete the non–word repetition task; six others did not complete the task due to experimenter error. CI = confidence interval.

The talkers who recorded the nonwords spoke the Standard American English dialect. Recordings were made in a sound booth, and the first author provided a model of each word for the talkers to repeat. One token of each word from each talker was extracted, with 50 ms of silence before and after the production. These extracted productions were then normalized for root-mean-square amplitude.

Previous studies have found that talker variability can facilitate production accuracy in young children (e.g., Plante et al., 2011; Richtsmeier et al., 2011), although the facilitation may be limited under some circumstances (e.g., Richtsmeier & Goffman, 2017; Zamuner et al., 2017). Furthermore, this is the first study to compare exposures with and without talker variability in a within-subject design. Based on the existing literature, we predicted a modest benefit from talker variability to both production accuracy and referent identification. No predictions were made regarding the interaction of talker variability and input frequency.

A colorful, hand-drawn picture of a make-believe animal accompanied each nonword (Ohala, 1999) to enhance semantic learning (Heisler, Goffman, & Younger, 2010). Eight different animals were used, one for each nonword, across eight different lists. The eight lists allowed for the input frequency and talker variability conditions to be counterbalanced in terms of animals and nonwords. Regarding animals, the animals were not balanced for semantic features, but each make-believe animal was assigned to a different nonword in each list. Thus, it is unlikely that the features of any make-believe animal would skew the results. Regarding the balance of nonwords, each nonword appeared in each of the eight conditions created by crossing input frequency and talker variability. Thus, the different phonotactic probabilities of the word-medial consonant sequences (see Table 2) contributed equally to each of the input frequency and talker variability conditions.

Table 2. Phonological properties of the word-medial consonant sequences (CC).

Word	CC	CC biphone probability
gɪsnək	sn	.0003
bozjəm	zj	.0000
tʌvʧəp	vʧ	.0000
sabləf	bl	.0015
fugdən	gd	.0001
mæfpəg	fp	.0000
niʃkət	ʃk	.0000
pɛmtəs	mt	.0002

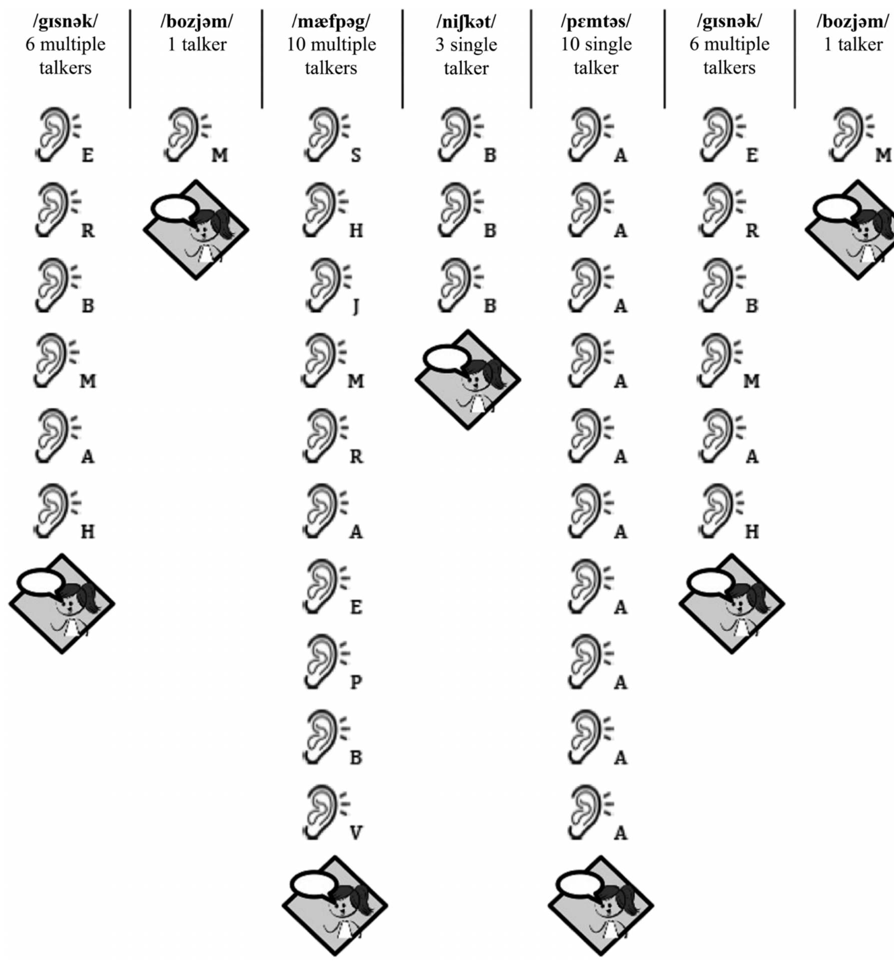
Procedure

To facilitate the description of the procedure, we use the following terminology. *Block* refers to separate sections of the experiment with unique purposes. The experiment comprised five blocks: test, ABX discrimination, referent identification, posttest, and a non–word repetition task based on Dollaghan and Campbell (1998). The test and posttest had additional structure: *Trial* refers to an opportunity to produce a nonword. There were three trials per nonword during the test and two trials per nonword during the posttest, or five total trials per nonword. Within each trial, the number of perceptual *exposures* varied depending on the target word's input frequency. Thus, a trial might include one, three, six, or 10 exposures prior to the child's opportunity to produce the word. During the posttest, the number of exposures for all nonwords was one.

Children were brought in by their parents for a single experimental session in a quiet room. Children sat at a child-sized table with an all-in-one computer in the center, speakers on both sides, and a mouse in front. Speaker volume was set to a comfortable level and was consistent across children. Presentation of the experiment was controlled by Paradigm computer software (Paradigm, 2015). The experimenter sat to the left of the child and provided directions and reinforcement. The child's parents observed the experiment in the room. Parents did not help their child with the experiment, but if their child became unhappy or distracted, parents occasionally provided encouragement.

Before starting the experiment, the experimenter explained to the child that he or she would play a game involving a set of make-believe animals. Children were familiarized with the trial structure in an initial training that featured the real English words *ball, kitty,* and *donut*. Four training trials were completed—*ball* was repeated twice—and the experimenter provided feedback throughout to support the child's understanding of the procedure. Experimental trials proceeded as shown in Figure 2. Visual support for the trial's structure included the make-believe animal at the top left of the screen, vertical blue boxes at the bottom of the screen for each exposure, plus an additional blue box for the child's production. As a sound file

Figure 1. An illustration of seven trials (out of 24) of the test condition for List 1. Each ear represents an exposure, whereas each child represents an opportunity for the child to produce the nonword. The letter next to the ears represents a unique talker. For example, /gɪsnək/ was heard six times and was presented by six different talkers.

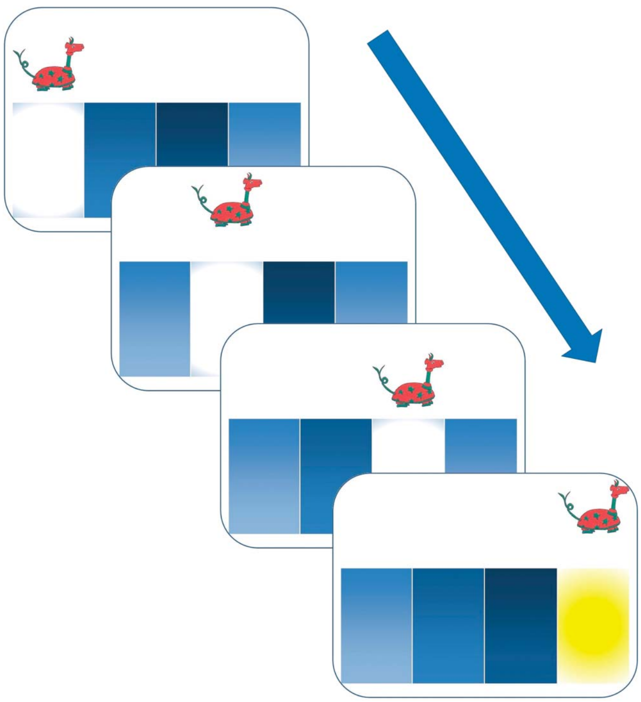


for each exposure played, the blue bar directly below the animal turned white. The make-believe animal then moved from left to right across the screen for each new exposure. When the animal advanced to the final box, the box turned yellow, prompting the child to produce the target nonword. If the child did not immediately produce the nonword, the experimenter would ask, "What was the animal's name?" If the child did not respond, the experimenter moved to the next trial. The child could control the pace of the exposures by touching the computer screen or clicking the mouse to advance to the next exposure. The experimenter could also move the experiment forward if the child preferred to listen passively. The experimenter controlled when a new trial began. Trials proceeded in a predetermined, pseudorandom order.

Following the test, the second experimental block was an ABX discrimination task in which the child heard a minimal pair of real English words that differed by one consonantal feature. An example of this would be /bæd/ and /dæd/, where the first consonants, /b/ and /d/, differ only in the feature of place of articulation (see the Appendix for the full list of items). Each word was played with an appropriate picture, and then one of the words was played once more with both pictures on the screen. The child was told to point to the picture he or she had heard. The purpose of this task was to determine if the child could distinguish the 16 sounds used in the nonword consonant sequences from other easily confusable sounds. However, in the analysis below, we only examine how a child's overall performance on the discrimination task relates to their production accuracy and referent identification. A more fine-grained analysis of the discrimination data is planned for a future analysis that will combine the results of multiple experiments.

The third block examined word learning via a referent identification task. Children completed four word-learning probes. For each probe, an array of three make-believe animals appeared on the computer screen, arranged horizontally. The experimenter asked which animal was associated

Figure 2. This figure depicts a trial from the test block with an input frequency of three. The first three boxes correspond to the three exposures to the target nonword. Those boxes turned white as the child heard the target nonword. The rightmost box turned yellow to indicate that it was the child's turn to produce the target nonword.



with a name that the experimenter produced. For example, the experimenter might ask, "Which animal is /pɛmtəs/?" The child would then point to the animal that he or she believed was associated with /pɛmtəs/. If the child did not point to an animal, the experimenter prompted a second time, for example, saying, "Point to the one you think is /pɛmtəs/." If the child did not point to an animal following the second prompt, the experimenter would say, "Point to the one you think it is." No child required more than two prompts to respond.

The fourth block consisted of a production posttest in which all nonwords had an input frequency of one. The posttest structure is illustrated in Figure 3 below. The posttest trial order was also pseudorandomized.

In the fifth and final block, children produced the 16 nonwords from Dollaghan and Campbell (1998). The nonwords increased in length from one to four syllables. The task is thought to tap phonological short-term memory. The results of this task were later entered into a correlation analysis along with the dependent measures of the main experiment, overall performance on the discrimination task, and standard scores from the GFTA-2.

The experiment ended immediately if a child communicated an unwillingness to continue (as was the case for six children, described above). Every child was given a small prize when the experiment ended, and families were compensated with $20 for their time.

Five productions of each experimental nonword were collected, so it was possible to investigate the influence of input frequency over several productions, thereby providing some insight into production practice. Production practice was not manipulated in the experiment but was consistent for all nonwords and for most children (n = 26). Contrary to the instructions, however, seven children repeated the nonwords during the familiarization and therefore gained unanticipated production practice. For instance, if /pɛmtəs/ was assigned to the input frequency level of 3, a child who repeated all familiarization items would have produced /pɛmtəs/ 14 times, whereas a child who completed the study as instructed would have produced /pɛmtəs/ just five times. We consider this extra production practice to be a natural response by children to the task demands, so we include their productions in the analyses below. We note, however, that we obtained equivalent results when data from these children were excluded from the analyses.

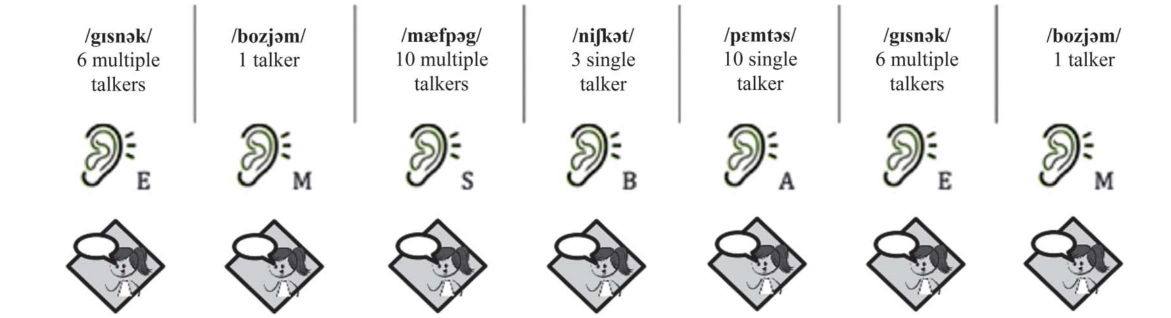
Analysis

Children's productions were transcribed phonetically by the second author and another graduate research assistant. Transcriptions of the four consonants for each nonword were then scored on a 3-point scale (Edwards, Beckman, & Munson, 2004; Richtsmeier et al., 2011). A score of 3 indicated an optimal production in terms of voicing, place of articulation, and manner of articulation. A score of 2 indicated that two of three features were produced accurately. A score of 1 indicated that two or more features were incorrectly produced but an attempt was made at producing the consonant. A score of 0 indicated that the consonant was deleted. Each nonword production was given a score out of 12 possible points.

As an example of the scoring system, if the target nonword was /gɪsnək/ and the child said [gizkə], then the child would receive 3 points for a correct production of /g/; 2 points for an incorrect production of /s/ that differed in voicing; 1 point for an incorrect production of /n/ that differed in place, manner, and voicing; and 0 points for no attempt at the word-final /k/. These scores would be totaled, and the child would receive 6 points for his or her production. It is possible that in the case of [gizkə], /n/ was omitted and the final /k/ was moved to the medial consonant sequence. However, this possibility did not factor into the scoring. All nonwords were scored as described above even if it were possible that movement, metathesis, or epenthesis had taken place.

A reliability analysis was conducted by having an undergraduate research assistant transcribe approximately 25% of the data—productions from nine participants—independently. Agreement was determined by comparing the scores calculated by each transcriber. There was 78% agreement overall, which we considered to be low. Additional reliability analyses were conducted on approximately 30% of the data. The first author independently transcribed data from four participants, and an undergraduate research

Figure 3. An illustration of seven trials (out of 16) of the posttest condition for List 1. In the posttest, each nonword was heard only once before being produced, regardless of the test conditions that are listed immediately below the nonword. For reference, the conditions from the test block are listed below each nonword. Each ear represents a perceptual exposure, whereas each child represents an opportunity for the child to produce the nonword. The initials next to the ears represent different talkers.



assistant transcribed data from six other participants. Agreement between the original transcriptions and the first author was 84%; between the original transcriptions and the undergraduate transcriber was 85%. We considered this scoring agreement to be acceptable. Note that only the original transcriptions made by the second author and the graduate research assistant were used in the analyses.

Regarding the referent identification task, a score of 1 was given for each correct match and a score of 0 for an incorrect match. Four different target animals (that is, the animals representing the correct answer for a given probe) were used for these word-learning probes, and the targets always had an input frequency level of 3, 6, or 10. For the first probe, children heard the target animal produced by a single talker, and the competitors were other single talker animals with different frequencies. For the second probe, the target animal was produced by multiple talkers, as were the competitors, and each option had a different input frequency. The third and fourth target animals compared learning in the single and multiple talker conditions. The third target animal was a multiple talker word, and it was paired with a competitor animal with the same input frequency from the single talker condition and another competitor that had a frequency of one. The fourth target animal was a single talker word with a different input frequency than the third target animal but with an equivalent set of competitors. Input frequencies of the third and fourth target animals were counterbalanced. Using this setup, the results of the four probes were combined and analyzed in terms of the questions set out in Table 3.

Results

Production Accuracy

Effects of all variables on mean production accuracy can be seen in Figure 4. Talker variability was only relevant to the 3, 6, and 10 levels of the input frequency factor, so to apply a parametric design, we conducted an analysis of variance (ANOVA) limiting input frequency to the values (3, 6, 10). Sphericity violations were found, so the results below are presented with Huynh–Feldt corrections.

Production accuracy scores were examined in a 2 × 2 × 3 design (Test Condition [test, posttest] × Talker Condition [single talker, multiple talkers] × Input Frequency [3, 6, 10]). Partial data were missing for four participants. There was a significant effect of test–posttest, $F(1, 28) = 6.95$, $MSE = 5.69$, $p = .014$, $\eta_p^2 = .20$. There was also a significant interaction of test–posttest and input frequency, $F(2, 56) = 6.45$, $MSE = 3.18$, $p = .003$, $\eta_p^2 = .187$. No significant effect of talker variability or a significant interaction with talker variability was found (all *p* values greater than .10).

Because input frequency appeared in a significant interaction in the first ANOVA, but only three of four input frequency conditions were included, a second 2 × 4 ANOVA (Test Condition [test, posttest] × Input Frequency [1, 3, 6, 10]) was conducted to better understand the interaction. A significant main effect of test–posttest was again observed, $F(1, 32) = 12.30$, $MSE = 8.97$, $p = .001$, $\eta_p^2 = .28$, with children performing more accurately in the posttest ($M = 10.90$) than in the test ($M = 10.53$). A main effect of input frequency was not significant, $F(2.50, 79.80) = 1.61$, $MSE = 2.77$, $p = .201$, $\eta_p^2 = .05$, but a near-significant trend was found in the Test × Frequency interaction, $F(2.03, 65.07) = 3.06$, $MSE = 2.35$, $p = .053$, $\eta_p^2 = .09$. Simple effects were subsequently analyzed. A significant effect of frequency was found in the test condition, $F(2.24, 76.93) = 3.34$, $MSE = 3.23$, $p = .028$, $\eta_p^2 = .10$, but not in the posttest condition, $F(2.22, 70.93) = 1.00$, $MSE = 1.41$, $p = .384$, $\eta_p^2 = .03$. As seen in Figure 5, accuracy during the test was highest for nonwords with Input Frequency 3. Post hoc contrasts were conducted to determine whether Input Frequency 3 resulted in significantly greater accuracy than the other input frequencies. Accuracy for Input Frequency 3 was significantly greater than Input Frequency 1, $F(1, 32) = 5.52$, $MSE = 13.90$, $p = .025$, and Input Frequency 6, $F(1, 32) = 5.46$, $MSE = 6.30$, $p = .026$, but not Input Frequency 10, $F(1, 32) = 1.00$, $MSE = 1.52$, $p = .324$.

Table 3. Six questions evaluated using data from the referent identification task.

Question	Mean accuracy	*SD*	*t*	*p*
1. Did children respond significantly above chance overall to all four questions?	52.3	28%	3.92	.000*
2. Did children respond significantly above chance when the target animal was Input Frequency 3?	50.0	50.8	1.89	.068
3. Did children respond significantly above chance when the target animal was Input Frequency 6?	50.0	45.8	2.10	.044
4. Did children respond significantly above chance when the target animal was Input Frequency 10?	57.8	42.3	3.32	.002*
5. Did children respond significantly above chance when the target animal was produced by multiple talkers?	48.4	37.0	2.36	.025
6. Did children respond significantly above chance when the target animal was produced by a single talker?	56.3	37.6	3.50	.001*

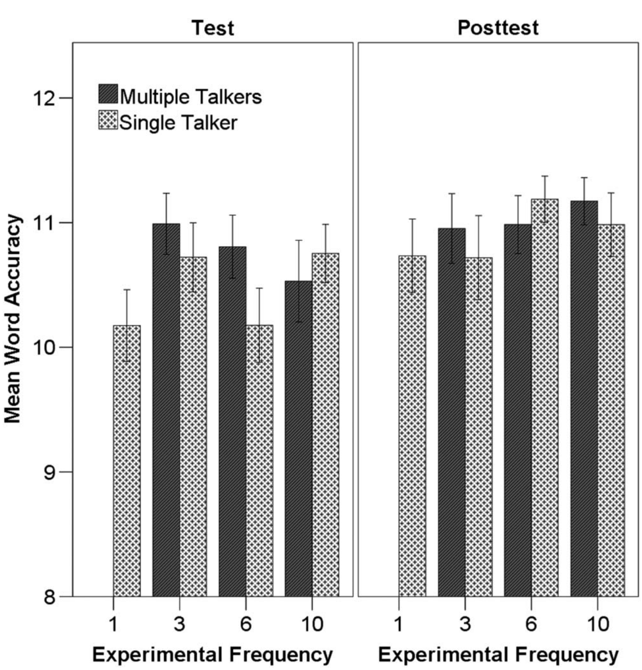
Note. Accuracy values significantly above chance were identified following a Šidák correction. One child was unable to complete the word-learning probes due to a computer error.

*$p < .008$.

Referent Identification

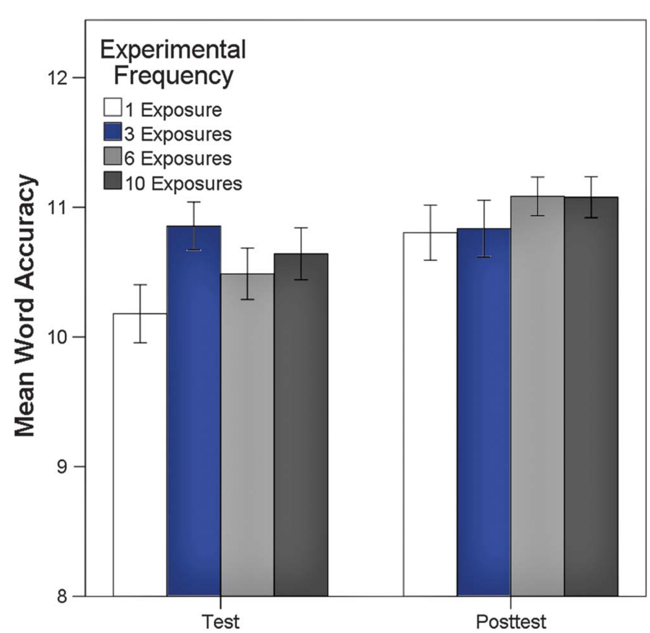
Results of the referent identification analysis are presented in Table 3. First, we considered whether or not children identified the referent of a word form with greater accuracy than that expected by chance (Question 1). Because each of the four word-learning probes involved three competitors, chance was set at 33.3%. Three analyses were then conducted regarding the benefits of Input Frequencies 3, 6, and 10 (Questions 2–4, respectively), and two analyses were conducted related to the multiple talker and single talker conditions (Questions 5 and 6).

Figure 4. Mean word accuracy out of 12 (*y*-axis) by input frequency (*x*-axis) and talker variability (multiple talkers represented by dark gray bars and single talker represented by light gray bars). Test data are shown in the left panel, and posttest data are shown in the right panel.



Because we completed six separate *t* tests, we applied a Šidák correction and adjusted α values to .008. On Question 1, we found that children were above chance in general (M = 52%), suggesting that they understood the task and were able to identify the make-believe animals when given their names. Children were also above chance when identifying nonwords with Input Frequency 10 (M = 58%) and nonwords that were produced by a single talker (M = 56%). However, the accuracy for all six questions was within 10 percentage points (range: 48.4–57.8%). We therefore conducted two ANOVAs to directly compare learning based on the different experimental manipulations. There was no significant difference in correct identification of targets with different input frequencies, $F(2, 62) = .34$, $MSE = .07$, $p = .716$. Similarly, participants did not correctly

Figure 5. Mean word accuracy (*y*-axis) broken down by test condition (*x*-axis) and input frequency (shading of bars).



identify more single talker targets compared with multiple talker targets, $F(1, 31) = 1.70$, $MSE = .39$, $p = .201$. In sum, although children were generally able to identify target animals with greater-than-chance accuracy, semantic learning may have been driven by general exposure rather than by manipulation of the experimental variables.

Pearson correlations were conducted to determine whether children's performance in the referent identification task related to performance in the production accuracy task or to any of the descriptive measures (GFTA-2 scores, auditory discrimination, and non–word repetition). Performance in the referent identification task was operationalized as each child's overall accuracy in the word learning probe. Performance in the production accuracy task was operationalized as the difference between accuracy for Input Frequencies 3 and 1. A total of 21 different correlation tests were conducted, so α was adjusted by Bonferroni correction to .002. There was a near-significant correlation between age and non–word repetition accuracy ($r = .56$, $p = .004$); older children tended to produce nonwords more accurately. All other correlations had $p \geq .02$.

General Discussion

In this study, children were tested on their ability to produce nonwords familiarized as the names of make-believe animals and to identify the make-believe animals when given their names. The nonwords were familiarized using four different input frequencies—one, three, six, and 10 exposures—and with or without talker variability across exposures. Children produced those nonwords three times in the test block, two more times in the posttest block, and identified corresponding referents for four of the nonwords in the intermediary.

Regarding production accuracy, there was an effect of input frequency observed during the test, and children were most accurate when producing nonwords that they had heard three times. However, the benefits of input frequency were not observable in the posttest, although children were significantly more accurate than in the test. No effects of talker variability were observed throughout. Regarding referent identification, children performed above chance. Nevertheless, ANOVAs did not reveal differences in identification based on the experimental manipulations. We now consider these results as they relate to the two questions raised in the introduction.

Is More Always Better?

In the introduction of this article, we observed that frequency effects typically correspond to a high-frequency benefit, leading to the hypothesis that "more is always better." In contrast to that prediction, children's production accuracy was numerically highest for nonwords that they heard three times during the test. Other researchers have reported that three is unique in supporting learning (Gerken & Bollt, 2008; Xu & Tenenbaum, 2007).

Although speculative, this finding that more is not always better may inform future clinical work, which can be conceptualized in terms of relatively short sessions or *teaching episodes* (Warren et al., 2007). In fact, 30-min meetings with clients is the most common meeting length for speech therapy in schools (Gillam, Baker, & Williams, 2012). Thus, speech-language pathologists must find ways to maximize that time by finding the ideal frequency or ideal dose within the context of a teaching episode. A study by Storkel et al. (2017) provides an evidence-based determination of the ideal dose for facilitating word learning during interactive book reading. The authors sought to identify the adequate teaching intensity, or dose, for new word learning. Based on groups of children receiving 12, 24, 36, and 48 exposures, a total of 36 exposures was adequate for learning to be observed, but 48 exposures did not lead to further improvement. Thus, Storkel et al. also provide compelling evidence that more is not always better. Future research is needed to identify the ideal input frequency dose for facilitating production accuracy in populations with delayed or disordered speech.

It is also worth considering the possibility that the ideal dose depends on the learning target. This can be seen when comparing the present finding of an optimal dose of three for production accuracy compared with Storkel et al.'s (2017) optimal dose of 36 for word learning. Even within the present results, three perceptual exposures appeared to be optimal for supporting production accuracy, but 10 exposures were necessary for participants to match a referent to its form at a level greater than chance. Although preliminary, we suggest that a greater dose may be necessary when targeting word learning compared with production accuracy. Thinking more generally about therapy, children may require more or less practice depending on the learning target. Future clinical work meant to establish recommendations for dose should consider various learning targets.

How Do the Effects of Input Frequency Play Out Across Repeated Productions?

The results contribute to a growing list of Input Frequency × Production Practice interactions in speech development. For example, Richtsmeier and Goffman (2017) observed that children only benefited from a high variability input for the first few productions they made of a novel word. Here, we observed an immediate benefit of input frequency during the test, but those benefits did not persist into posttest productions.

The results are also consistent with recent observations that production may inhibit word learning. Zamuner et al. (2017) found that practice producing the name of a make-believe animal resulted in less consistent recognition in a forced-choice test. Children in that study were also less likely to freely recall the animal names that they produced in comparison to the animal names that they listened to. Unfortunately, a weakness of our study is that it does not allow us to determine how production practice affected word learning. To observe the interaction of input frequency and production practice in word learning, future research should include a referent identification probe following various production frequencies.

Another weakness of this study is that not all children performed the task in the same way, and seven children gained additional production practice by repeating the exposure words. Although future studies may attempt to control for this difference by requiring all children to listen during the familiarization, or by removing from the analyses any children who do not perform the task as instructed, we believe that these approaches are premature. Our intuition is that repeating a new word immediately is a natural response when learning a new word, at least for some children. We therefore believe that researchers should treat this kind of production practice as a type of individual variability worthy of future study.

Regarding talker variability, the findings do not suggest that talker variability is instrumental to learning from the input. In comparison to previous studies (Plante et al., 2011; Richtsmeier et al., 2009), talker variability did not improve production accuracy beyond the raw increase in input frequency in the single talker condition. Because we used a within-subject design, our results represent one of the strongest tests to date of talker variability. Nevertheless, there are still reasons to believe that talker variability can be facilitative. For example, two studies of adult Japanese learners of English found that talker variability could be part of a robust perceptual training regimen for learning the English /r/-/l/ contrast (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). Those studies included training across many weeks, whereas this study is limited to learning within an hour. Thus, talker variability may provide a relatively small benefit to perceptual learning, but that benefit may accrue over time. From a theoretical perspective, there are also compelling arguments that talker variability may help establish the boundaries of perceptual categories (Pierrehumbert, 2003; Richtsmeier et al., 2011). Thus, future research on talker variability may be most fruitful if it focuses on newly learned or uncertain perceptual categories.

Finally, using a test–posttest structure, the study examined the effects of input frequency and production practice over time. From the perspective of word learning, the test corresponds to a learning stage that Storkel (2015) refers to as "learning from input (p. 2)" whereas the posttest corresponds to a stage called *memory evolution,* or a gap in learning or intervention. As reviewed by Storkel, these two stages are supported by separable neurocognitive processes, and in terms of language delays, each stage may be uniquely vulnerable to errors. Storkel's analyses suggest that repeated cycles of learning and memory evolution provide an important perspective as to whether a particular child's word-learning difficulties reflect impaired learning from the input or memory evolution. Although this study does not allow for an analysis of learning cycles, future studies of the interaction of input frequency and production practice may benefit from this methodology.

Acknowledgments

The authors wish to thank Anna Bostian, Shie Kantor, and Kelsey Mueggenborg for help in managing the participants.

References

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language, 42*(2), 239–273. https://doi.org/10.1017/S030500091400049X

Ambridge, B., Pine, J. M., Rowland, C. F., Chang, F., & Bidgood, A. (2013). The retreat from overgeneralization in child language acquisition: Word learning, morphology, and verb argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(1), 47–62.

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. https://doi.org/10.1016/j.jml.2015.10.008

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977–985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299–2310.

Capone, N. C., & McGregor, K. K. (2005). The effect of semantic representation on toddlers' word retrieval. *Journal of Speech, Language, and Hearing Research, 48*(6), 1468–1480. https://doi.org/10.1044/1092-4388(2005/102)

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*(5), 1136–1146.

Edeal, D. M., & Gildersleeve-Neumann, C. E. (2011). The importance of production frequency in therapy for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 20*(2), 95–110. https://doi.org/10.1044/1058-0360(2011/09-0005)

Edwards, J., & Beckman, M. E. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning and Development, 4*(2), 122–156. https://doi.org/10.1080/15475440801922115

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur–Bates Communicative Development Inventories*. Baltimore, MD: Brookes.

Gerken, L., & Bollt, A. (2008). Three exemplars allow at least some linguistic generalizations: Implications for generalization mechanisms and constraints. *Language Learning and Development, 4*(3), 228–248.

Gershkoff-Stowe, L. (2002). Object naming, vocabulary growth, and the development of word retrieval abilities. *Journal of*

Memory and Language, 46(4), 665–687. https://doi.org/10.1006/jmla.2001.2830
Gershkoff-Stowe, L., Connell, B., & Smith, L. (2006). Priming overgeneralizations in two- and four-year-old children. *Journal of Child Language, 33*(3), 461–486.
Gillam, R., Baker, E., & Williams, A. L. (2012). *How much is enough? Dosage in child language intervention*. Paper presented at the American Speech-Language-Hearing Association Convention, Atlanta, GA. Retrieved from http://www.asha.org/Events/convention/handouts/2012/1465-How-Much-Is-Enough-Dosage-in-Child-Language-Intervention/
Gladfelter, A., & Goffman, L. (2018). Semantic richness and word learning in children with autism spectrum disorder. *Developmental Science, 21*(2), e12543. https://doi.org/10.1111/desc.12543
Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2)*. San Antonio, TX: Pearson.
Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*(3), 515–531.
Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102*(3), 594–621.
Heisler, L., Goffman, L., & Younger, B. (2010). Lexical and articulatory interactions in children's language production. *Developmental Science, 13*(5), 722–730. https://doi.org/10.1111/j.1467-7687.2009.00930.x
Hewlett, N. (1990). Processes of development and production. In P. Grunwell (Ed.), *Developmental speech disorders* (pp. 15–30). Edinburgh, United Kingdom: Churchill Livingstone.
Hickok, G. (2014). Toward an integrated psycholinguistic, neurolinguistic, sensorimotor framework for speech production. *Language, Cognition and Neuroscience, 29*(1), 52–59. https://doi.org/10.1080/01690965.2013.852907
Hodson, B., & Paden, E. (1991). *A phonological approach to remediation: Targeting intelligible speech*. Austin, TX: Pro-Ed.
Kan, P. F., & Sadagopan, N. (2015). Speech practice effects on bilingual children's fast mapping performance. *Seminars in Speech and Language, 36*(2), 109–119. https://doi.org/10.1055/s-0035-1549106
Kaushanskaya, M., & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes, 26*(1), 121–148.
Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 11850–11857. https://doi.org/10.1073/pnas.97.22.11850
Maas, E., Butalla, C. E., & Farinella, K. A. (2012). Feedback frequency in treatment for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 21*(3), 239–257. https://doi.org/10.1044/1058-0360(2012/11-0119)
Maas, E., & Farinella, K. A. (2012). Random versus blocked practice in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55*(2), 561–578. https://doi.org/10.1044/1092-4388(2011/11-0120)
Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)
Masdottir, T., & Stokes, S. F. (2016). Influence of consonant frequency on Icelandic-speaking children's speech acquisition. *International Journal of Speech-Language Pathology, 18*(2), 111–121. https://doi.org/10.3109/17549507.2015.1060525
McAllister Byun, T., Inkelas, S., & Rose, Y. (2016). The A-map model: Articulatory reliability in child-specific phonology. *Language, 92*(1), 141–178.
McLeod, S. (2016). Speech sound acquisition. In J. E. Bernthal, N. W. Bankson, & P. J. Flipsen (Eds.), *Articulation and phonological disorders: Speech sound disorders in children* (Vol. 8, pp. 49–92). Boston, MA: Pearson.
Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research, 44*(4), 778–792.
Ohala, D. K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders, 32*(6), 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*(3), 539–566. https://doi.org/10.1017/S0305000912000074
Paradigm. (2015). Paradigm (Version 2.4) [Computer software]. Lawrence, KS: Perception Research Systems. Retrieved from http://www.paradigmexperiments.com/
Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46,* 115–154. https://doi.org/10.1177/00238309030460020501
Plante, E., Bahl, M., Vance, R., & Gerken, L. (2011). Beyond phonotactic frequency: Presentation frequency effects word productions in specific language impairment. *Journal of Communication Disorders, 44*(1), 91–102. https://doi.org/10.1016/j.jcomdis.2010.07.005
Rescorla, L., & Ratner, N. B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech and Hearing Research, 39*(1), 153–165.
Richtsmeier, P. T., Gerken, L., Goffman, L., & Hogan, T. (2009). Statistical frequency in perception affects children's lexical production. *Cognition, 111*(3), 372–377. https://doi.org/10.1016/j.cognition.2009.02.009
Richtsmeier, P. T., Gerken, L., & Ohala, D. (2011). Contributions of phonetic token variability and word-type frequency to phonological representations. *Journal of Child Language, 38*(5), 951–978. https://doi.org/10.1017/S0305000910000371
Richtsmeier, P. T., & Goffman, L. (2017). Perceptual statistical learning over one week in child speech production. *Journal of Communication Disorders, 68,* 70–80. https://doi.org/10.1016/j.jcomdis.2017.06.004
Robb, M. P., & Bleile, K. M. (1994). Consonant inventories of young children from 8 to 25 months. *Clinical Linguistics & Phonetics, 8*(4), 295–320. https://doi.org/10.3109/02699209408985314
Rvachew, S., & Brosseau-Lapré, F. (2015). A randomized trial of twelve-week interventions for the treatment of developmental phonological disorder in Francophone children. *American Journal of Speech-Language Pathology, 24,* 637–658. https://doi.org/10.1044/2015_AJSLP-14-0056
Smit, A. B. (1993). Phonologic error distributions in the Iowa–Nebraska Articulation Norms Project: Word-initial consonant clusters. *Journal of Speech and Hearing Research, 36*(5), 931–947.
Stemberger, J. P., & Bernhardt, B. H. (1999). The emergence of faithfulness. In B. MacWhinney (Ed.), *The emergence of language* (pp. 417–446). Mahwah, NJ: Erlbaum.
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* (Vol. 8, pp. 25–52). Baltimore, MD: Brookes.
Stoel-Gammon, C., & Dale, P. (1988). *Aspects of phonological development of linguistically precocious children*. Paper presented at the Child Phonology Conference, University of Illinois, Champaign–Urbana.

Stokes, S. F., & Surendran, D. (2005). Articulatory complexity, ambient frequency, and functional load as predictors of consonant development in children. *Journal of Speech, Language, and Hearing Research, 48*(3), 577–591.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*(6), 1321–1337.

Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research, 46*(6), 1312–1323.

Storkel, H. L. (2015). Learning from input and memory evolution: Points of vulnerability on a pathway to mastery in word learning. *International Journal of Speech-Language Pathology, 17*(1), 1–12.

Storkel, H. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying an adequate intensity and variation in treatment response. *Language, Speech, and Hearing Services in Schools, 48*(1), 16–30. https://doi.org/10.1044/2016_LSHSS-16-0014

Tatsumi, T., Ambridge, B., & Pine, J. M. (2017). Disentangling effects of input frequency and morphophonological complexity on children's acquisition of verb inflection: An elicited production study of Japanese. *Cognitive Science, 42*(S2), 555–577.

Templin, M. C. (1957). *Certain language skills in children; their development and interrelationships*. Minneapolis, MN: University of Minnesota Press.

Vaden, K. I., Halpin, H., & Hickok, G. S. (2009). *Irvine phonotactic online dictionary*. (Version 2.0) [Data file]. Retrieved from http://www.iphod.com

Van Riper, C., & Erickson, R. L. (1996). *Speech correction: An introduction to speech pathology and audiology* (9th ed.). Boston, MA: Allyn & Bacon.

Vihman, M. M. (1992). Early syllables and the construction of phonology. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 393–422). Timonium, MD: York Press.

Vihman, M. M., DePaolis, R. A., & Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning, 64*(2), 121–140.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Developmental Disabilities Research Reviews, 13*(1), 70–77. https://doi.org/10.1002/mrdd.20139

Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10*(3), 288–297. https://doi.org/10.1111/j.1467-7687.2007.00590.x

Zamuner, T. S., Gerken, L., & Hammond, M. (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language, 31*(3), 515–536.

Zamuner, T. S., Strahm, S., Morin-Lessard, E., & Page, M. P. (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

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*.

Appendix

Auditory Discrimination Word Pairs

Word 1	Word 2	Contrast
dot	tot	d/t
buy	guy	b/g
shock	sock	ʃ/s
cub	cup	b/p
hiss	hit	s/t
peas	peace	z/s
dad	bad	d/b
back	bag	k/g
leave	leaf	v/f
care	pair	k/p
lake	wake	l/w
fan	pan	f/p
pair	bear	p/b
sew	toe	s/t
walk	rock	w/r
tough	cuff	t/k
bride	bribe	d/b
yuck	luck	j/l
leash	leech	ʃ/tʃ
bed	beg	d/g
van	fan	v/f
choke	joke	tʃ/dʒ
rug	rub	g/b
kneel	near	l/r
win	bin	w/b
red	lead	r/l
nine	dine	n/d
chip	ship	tʃ/ʃ
cuff	cup	f/p
zoo	Sue	z/s
top	pop	t/p
bed	bet	d/t
gash	gas	ʃ/s
dig	gig	d/g
gap	cap	g/k