

Order Effects in the Perception and Production of New Words

Peter T. Richtsmeier

Oklahoma State University

Michelle W. Moore

West Virginia University

Peter T. Richtsmeier, Department of Communication Sciences and Disorders, Oklahoma State University; Michelle W. Moore, Department of Communication Sciences and Disorders, West Virginia University.

Correspondence should be addressed to Peter T. Richtsmeier, Department of Communication Sciences and Disorders, Oklahoma State University, 042 Murray Hall, Stillwater, OK, 74078. Email: richtsm@okstate.edu

Conflict of Interest Statement: The authors report no conflicts of interest with regards to the funding, equipment, or supplies used to conduct this study. The authors report no financial or nonprofessional benefits outside of career advancement.

Abstract

Purpose: Perceptual learning and production practice are basic mechanisms that children depend on to acquire adult levels of speech accuracy. In this study, we examined perceptual learning and production practice as they contributed to changes in speech accuracy in three- and four-year-old children. Our primary focus was manipulating the order of perceptual learning and baseline production practice to better understand when and how these learning mechanisms interact.

Method: Sixty-five typically-developing children between the ages of three and four were included in the study. Children were asked to produce CVCCVC nonwords like /bozjəm/ and /tʌvtjəp/ that were described as the names of make-believe animals. All children completed two separate experimental blocks: a baseline block in which participants heard each nonword once and repeated it, and a test block in which the perceptual input frequency of each nonword varied between 1 and 10. Half of the participants completed a baseline-test order; half completed a test-baseline order.

Results: Greater accuracy was observed for nonwords produced in the second experimental block, reflecting a production practice effect. Perceptual learning resulted in greater accuracy during the test for nonwords that participants heard 3 or more times. However, perceptual learning did not carry over to baseline productions in the test-baseline design, suggesting that it reflects a kind of temporary priming. Finally, a post hoc analysis suggested that the size of the production practice effect depended on the age of acquisition of the consonants that comprised the nonwords.

Conclusions: The study provides new details about how perceptual learning and production practice interact with each other and with phonological aspects of the nonwords, resulting in

complex effects on speech accuracy and learning of form-referent pairs. These findings may ultimately help speech-language pathologists maximize their clients' improvement in therapy.

Keywords: child speech; speech production; speech perception; learning; consonant age of acquisition

In young children, learning the form and meaning of new words is a basic component of language development. What aspects of a child's experience facilitate that word learning? One aspect is perception, or the act of listening to words in the ambient speech. We refer to this as input-based perceptual learning, or *perceptual learning* for short. A second aspect is production, or the act of producing new words, referred to hereafter as *production practice*. Both perceptual learning and production practice result in learning, but recent research suggests that their effects are not always additive. Perception and production have produced divergent effects—or interactions—on word learning and speech production accuracy. In this paper, we review three such effects from the literature. We then present a study that builds on a previously reported interaction. The results reaffirm that perceptual learning and production practice can interact. Although we do not resolve the complexity of their relationship, it appears that the type of task and at least some phonological aspects of the words to be learned help explain why perceptual learning and production practice have divergent effects.

Perception-Production Interactions

Although perception and production support learning independently, emerging research examines cases where their effects diverge or interact. In a study with English-speaking adults, Kaushanskaya and Yoo (2011) were interested in how perceptual learning and production practice influence memory for the meanings of novel words with different levels of phonological familiarity. Production practice generally improves memory for newly learned words with well-known sound patterns (for example, Pritchard, Heron-Delaney, Malone, & MacLeod, 2019). Kaushanskaya and Yoo's first experiment replicated those benefits: For novel words like /tuf/ and /gef/, adult participants were able to recall English translations and identify the correct translation for novel words that they produced compared to novel words that they only heard

and practiced silently. In the second experiment, however, participants were asked to learn novel words like /tʏf/ and /χɛf/ that contained non-English sound patterns. Based on the same recall and identification tasks, participants learned more words that they only heard. Taking both experiments into account, production practice and perceptual learning did not support word learning equivalently but instead interacted, depending on the words to be learned. Novel words composed of familiar sound sequences benefited from production practice. Novel words composed of unfamiliar sound patterns instead benefited from perceptual learning. At least in the short term, when a learner has limited phonological knowledge, perceptual learning appears preferable for learning a word's meaning (see also, Baese-Berk & Samuel, 2016).

The effects of perceptual learning and production practice also diverged in a word learning study with children. Zamuner, Strahm, Morin-Lessard, and Page (2018) compared novel word learning via perception and production in four- to six-year-olds using eye tracking. Similar to Kaushanskaya and Yoo's (2011) first experiment, the novel words were composed of common sound sequences, for example, /zɛl/ and /mig/, and were paired with make-believe animals. Children's looks to the target animal were used to gauge learning. Children looked faster to the referents of words they heard compared to words they produced. A recall task completed after the eye-tracking experiment also favored perceptual learning—children recalled more heard-only words. Zamuner et al. conclude by building on Kaushankaya and Yoo's claim that production may heighten a mismatch between a learner's pre-existing knowledge and a word's form. They suggest that production practice hinders word learning when the participants have limited phonological knowledge, in general, and not just in cases of nonnative sounds. In contrast, perceptual learning appears to be facilitative in those same circumstances. Similar to adult second language learners, four- to six-year-olds may not see immediate benefits from production practice when learning the meaning of a new word.

Production practice does not need to be detrimental for it to diverge from perceptual learning, however. Richtsmeier and Good (2018) compared perceptual learning and production practice in a word learning study with children that focused primarily on production accuracy. During the initial test block, children were exposed to and produced eight novel two-syllable words like /boʒjəm/ and /tʌvtʃəp/. The test involved a perceptual manipulation of input frequency—children heard the novel words 1, 3, 6, or 10 times before each attempt at production. Children more accurately produced words they heard three or more times. However, the perceptual manipulations did not affect production accuracy in the posttest. Rather, children were more accurate for all words in the posttest, indicating they benefitted from production practice. In sum, Richtsmeier and Good found evidence for both perception- and production-based learning, but their effects interacted. Initial productions benefitted from perceptual learning, but later productions only showed an effect of production practice. Compared with results from the previous two studies, these results suggest that production practice may be more beneficial for improving speech accuracy than for helping a person learn form-referent pairings.

Present Study

The present study builds directly on the work from Richtsmeier and Good (2018). Those authors found effects of both perceptual learning and production practice, but no perceptual learning effects in the posttest. At least three possible explanations could be given for limited perceptual learning. The first is that perceptual learning may be ephemeral, and it largely reflects temporary priming or activation of pre-existing knowledge. By the time children moved to the posttest, the phonological knowledge primed by the perceptual manipulations may have dissipated. A second reason could be the act of production itself. The cognitive resources

required to produce a novel word appear to limit identification and recall of the referent (Kaushanskaya & Yoo, 2011; Zamuner et al., 2018). Perhaps production also limits or prevents what could have been learned via perception about a word's correct form. Finally, the combination of perceptual learning being both ephemeral and outweighed by production practice reflects a third possibility. Here we most directly test the second possibility, that is, that production practice limits perceptual learning. The present study addresses the following research question: Does initial production practice negatively affect the benefits of perceptual learning to speech production accuracy? Motivated in part by the findings from Kaunshanskaya and Yoo (2011) in which production practice and perceptual learning effects varied depending on the phonological composition of the words, an ancillary question will be considered in post hoc analyses: Do the phonological properties of the word learning items predict the perceptual learning and production practice effects?

The present study attempts to answer these questions using both new and previously published data. Regarding published data, Richtsmeier and Good (2018) used a test-posttest design in which the input frequencies of eight nonsense words varied during the test. Participants then produced the nonsense words again without the input frequency manipulation, providing a posttest baseline. The present study uses the same materials, but participants completed a pretest-test design in which baseline productions were collected first, and then input frequency manipulations followed. Thus, the only difference between the Richtsmeier and Good data and the new data here is the order of events: test-baseline versus baseline-test, respectively. The following hypothesis is based on the second explanation for Richtsmeier and Good's data: In the novel baseline-test condition, productions at baseline will prevent perceptual learning effects in the test. In contrast, the first and third explanations do not preclude a perceptual learning effect in the baseline-test condition. They are therefore

consistent with perceptual learning effects at test in both the test-baseline and baseline-test conditions.

Method

An overview of the relevant conditions of the study can be seen in Table 1 below. Some methodological details are only briefly described here. Page references are made to fuller descriptions in Richtsmeier and Good (2018).

Table 1 – A visualization of the study design, including how exposures and trials were distributed across the baseline and test conditions, as well as how the baseline and test conditions differed across the two order conditions.

Order Condition	First Block	Second Block
Test-Baseline Order (Data from Richtsmeier & Good, 2018)	3 trials, input frequency manipulation of exposures	2 trials, 1 exposure per word
Baseline-Test Order (Novel data)	2 trials, 1 exposure per word	3 trials, input frequency manipulation of exposures

Participants

In total, 77 children were recruited for the study, 41 of whom were reported on previously. All children were between the ages of 3;0 and 5;1 ($M = 3;11$), and they were recruited from the surrounding area of approximately 80,000 people through online social media and traditional media. Children's data were included in the analyses if they met criteria for typical development based on rule-in procedures. First, they completed the Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2; Goldman & Fristoe, 2000); only children receiving a

standard score of 85 or above were retained for the analysis. A hearing screening with pure tones in each ear of 1000, 2000, and 4000 Hz at 25 dB was administered to ensure that participants had normal hearing. Using a parent/guardian questionnaire that included questions about the children's developmental history, all participants were required to have no history of serious developmental or medical issues. Finally, only children acquiring English as their first and primary language were included. In the questionnaire, the parent/guardian of 18 children reported some exposure to a second language; for 2 children they reported exposure to a third language. Spanish was the most common other language ($n = 14$). No child was reported to receive more than three hours per week of exposure to another language, and for children with multilingual exposure, average exposure to another language each week was about 1.1 hours.

From the initial pool of 77, data were discarded from 12 children. Data from six children were removed because the study was not completed. Two more children failed both the hearing screening and a hearing test. One child received a score below 85 on the GFTA-2 ($SS = 79$), the parents of one child reported a developmental delay, and one child was a non-native speaker of English. One child's data were also lost due to experimenter error. All available data from the remaining 65 children, 35 female and 30 male, were included in the analyses. In addition to data from the rule-in procedures above, descriptive data were collected, although these data were not anticipated to relate to performance in the experiment, and they are not described in detail. Children completed an ABX auditory discrimination task in which they heard in succession two real words with picture referents, followed by both pictures and just one of the words; the task was to point to the picture corresponding to the word that they heard. Children also completed a nonword repetition task—with no perceptual manipulations or semantic referents—using nonwords from Dollaghan and Campbell (1998). Parental measures of children's gross and fine motor skill, as well as the family's socioeconomic status, were collected using 5-point scales, and

signatures from the assent forms were used to score name writing ability on a separate 5-point scale (Puranik, Schreiber, Estabrook, & O'Donnell, 2014). Summary data from these descriptive tasks, the rule-in procedures, and the results of the semantic probe are presented in Table 2 below. We note that some unanticipated differences between the two participant groups were observed. Those differences are considered in the General Discussion.

Table 2 - Mean scores, standard deviations, and confidence intervals for the descriptive measures collected from participants, with separate columns for the two order conditions.

	Test-Baseline Order			Baseline-Test Order			<i>p</i> ^a
	n	<i>M</i> (<i>SD</i>)	95% CI	n	<i>M</i> (<i>SD</i>)	95% CI	
Age in months	33	45.67 (6.06)	[43.60, 47.73]	32	48.94 (6.68)	[46.62, 51.25]	0.043*
GFTA-2 standard scores	33	112.72 (7.62)	[110.0, 115.45]	32	109.63 (8.57)	[106.51, 112.74]	0.128
Auditory discrimination	25	75% (17%)	[68%, 82%]	32	61% (20%)	[54%, 68%]	0.005**
Non-word repetition	24	75% (13%)	[69%, 80%]	26	65% (12%)	[60%, 70%]	0.013*
Name writing, 0-4 scale	24	1.67 (1.46)	[1.05, 2.28]	32	1.94 (1.41)	[1.42, 2.45]	0.488
Gross motor, 1-5 scale	21	3.43 (0.68)	[3.13, 3.73]	32	3.38 (0.75)	[3.1, 3.65]	0.793
Fine motor, 1-5 scale	21	3.14 (0.57)	[2.89, 3.4]	32	3.27 (0.65)	[3.03, 3.5]	0.484
SES, 1-5 scale	29	4.1 (0.62)	[3.87, 4.34]	32	3.72 (0.96)	[3.37, 4.07]	0.071†
Semantic probe accuracy	32	48% (30%)	[38%, 59%]	32	37% (26%)	[27%, 46%]	0.038*

Note. GFTA-2 = Goldman-Fristoe Test of Articulation-2, SES = Socioeconomic status, CI = Confidence Interval. ^aAll but one of the statistical comparisons were made using two-tailed student's *t* tests. The comparison for semantic probe accuracy was made using a binomial regression. - 2 log likelihood = 364.26, $\theta = -.52$. † $p < .10$. * $p < .05$. ** $p < .01$

Materials

During the experiment, children were tasked with learning eight CVCCVC nonwords (*items* hereafter) with initial stress: /pɛmtəs/, /nɪjkət/, /mæfpæg/, /fugdən/, /sabləf/, /tʌvtjəp/, bozjəm/, and /gɪsnək/. A full description of the inherent properties of these items, including the selection of consonants, statistics related to phonotactic probability, and neighborhood density can be found in Richtsmeier and Good (pp. 2858-9). A make-believe animal accompanied each item (Ohala, 1999), and participants were told that the items were the make-believe animals' names.

Each item was assigned an input frequency and a degree of talker variability, the two manipulations of perceptual learning. The input frequencies of 1, 3, 6, and 10 represent the number of perceptual exposures that a child heard before producing the item. Two items had a frequency of 1, two had a frequency of 3, and so on. For the two items with input frequencies of 3, 6, or 10, one item's exposures came from a single talker (i.e., one repeated sound file). Exposures of the other item came from multiple talkers, resulting in a total of seven within-subject conditions. Because only 1 talker is possible for an experimental frequency of 1, only the single talker condition exists for that input frequency. A fuller description of the talkers can be found in Richtsmeier and Good (pp. 2858-9).

To ensure that the perceptual learning manipulations were not confounded with inherent properties of the items, eight lists were created for each of the two order conditions (baseline-test condition and test-baseline condition), or sixteen total lists. Across lists, each item appeared in each of the within-subject conditions created by crossing four levels of input frequency and two levels of talker variability. As an additional step to create balance, the make-believe animal associated with an item also varied across the lists, with each item being associated with four different make-believe animals.

Procedure

Table 1 above provides an overview of the procedure. The primary manipulation in this study was the relative order of the baseline and test, a factor hereafter referred to as *order*. In Richtsmeier and Good's study, participants completed a test-posttest baseline design in which no perceptual learning manipulations were made during the posttest. In the pretest baseline-test order condition added here, no perceptual learning manipulations were made in the pretest baseline. We use the term *baseline* to indicate an experimental block in which there were no perceptual learning manipulations made and to help clarify that the pretest baseline block is identical to the posttest baseline—except that it occurred prior to the test. We use the term *test* to indicate an experimental block in which the perceptual manipulation occurs. Thus, the order factor compares the test-baseline condition from Richtsmeier and Good to a baseline-test condition. The temporal order or sequencing of blocks is also important, and the terms *first block* and *second block* refer to the sequencing of the blocks, regardless of order condition.

We use the following additional terminology: *Trial* refers to opportunities to produce one of the eight experimental items, and all children completed five trials per item over the course of the experiment—two trials during baseline and three trials during test—for a total of 40 trials overall. *Exposure* reflects input frequency within a trial. It refers to the number of times the child heard an experimental item before producing it. There was always 1 exposure during baseline, but exposures were 1, 3, 6, or 10 during test.

Parents brought their children to a laboratory on a university campus to complete the single session required for the study. Children first completed the GFTA-2. They then moved to a table with an all-in-one computer in the center, a mouse in front, and speakers on either side. The experimenter sat to the child's left and held a keyboard. Presentation of the experiment was controlled by Paradigm

computer software (Paradigm, 2015). Children were familiarized with the structure of the test in an initial training. The training mimicked the structure of the test (described below) but used the real words *kitty*, *donut*, and *ball*. Children were instructed to listen during the exposures to each word and then repeat what they heard after the last exposure. After the training, the experimenter explained that they would play a game with make-believe animals, and the child's task was to learn about the animals. At that point, the experiment's first block began—33 children completed the test, and 32 completed the baseline. All children then completed an auditory discrimination task and then a referent identification task. These tasks were followed by the experiment's second block, and children either completed the test or the baseline, whichever they had not completed earlier. The study ended after children completed a nonword repetition task (Dollaghan & Campbell, 1998).

During both baseline and test, visual and auditory cues were present to help direct children when to listen to the make-believe animal's names and when to produce them. As shown in Figure 1, during each trial, there were blue bars at the bottom of the screen. There was one bar for each exposure and the child's production. At the start of each trial, the make-believe animal appeared above the leftmost bar. The bar turned white, and the first exposure of the item played through the speakers. The picture of the animal then moved to the next bar to the right. When it reached the final blue bar, the bar turned yellow. For the first few trials, the experimenter prompted the child to produce the item once the rightmost bar turned yellow. If the child did not produce the item, the experimenter asked up to two times for a production. If the child still did not produce the item, the experimenter moved the experiment to the next trial. The intention was to have children only listen to each exposure, however 14 children repeated the items on 25% or more of the exposures during the test block. We return to these participants in the Results and General Discussion.

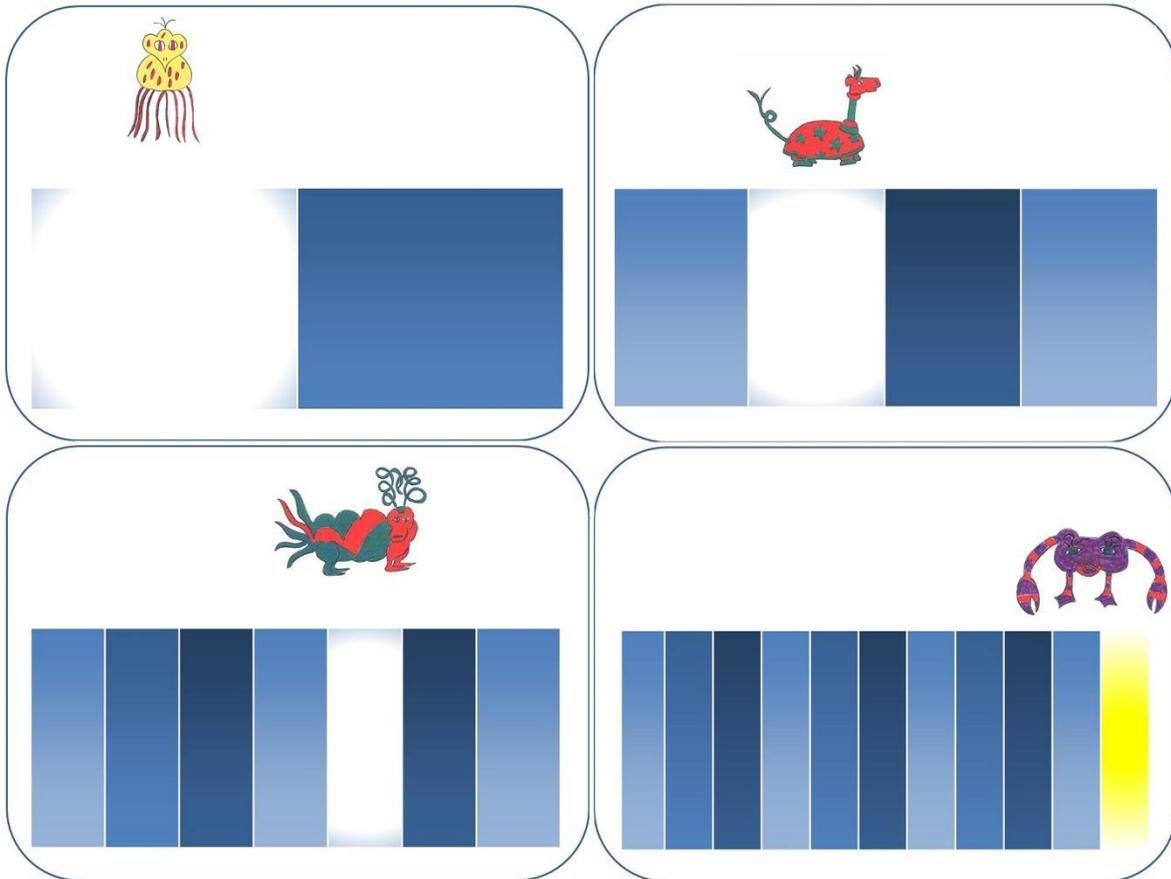


Figure 1 – Examples of the visual supports for the trial structure: The screen for the first and only exposure from the baseline and input frequency 1 conditions (top left); the second exposure from the input frequency 3 condition (top right); the fifth exposure from the input frequency 6 condition (bottom left); the production prompt from the input frequency 10 condition (bottom right).

The experiment automatically moved from one exposure to the next every 1.25 seconds. However, children were able to control the pace of the exposures, and as soon as the sound file for the nonword played, they could use the mouse or the touch screen to move to the next exposure. Movement from trial to trial and from block to block was controlled by the experimenter.

The experiment ended immediately if a child communicated an unwillingness to continue, as was the case for six children whose data were removed from analyses. Following the completion of the

experiment, every child was given a small prize, and families were compensated with \$20 for their time. These procedures were approved by the sponsoring university's institutional review board.

Analysis

Transcriptions of children's productions were made offline from video or audio recordings. Transcriptions were made using a broad transcription system by three research assistants as well as the first author. Production accuracy was calculated based on accuracy for the four consonants in each item using a 3-point system, with one point each for the consonant's voicing, place of articulation, and manner of articulation (Edwards, Beckman, & Munson, 2004). For example, a child who produced /mæfræg/ as [mæfræg] would receive 3 points for each consonant for a maximum possible score of 12. A child who produced /tʌvtʃəp/ as [tɛdʒəps] would receive 3 points for the /t/, 0 points for the omitted /v/, 2 points for /tʃ/ due to a change in voicing, and 1 point for the /p/ that was replaced by a consonant cluster. No score was given for productions that were not attempted (n = 129, or approximately 2 missing productions per participant).

Reliability analyses for the test-baseline data were previously reported by Richtsmeier and Good, and reliability averaged about 84%. For the new baseline-test data, a reliability analysis was completed by a research assistant who was unfamiliar with the purpose of the study and who re-transcribed productions from 13 of the 32 participants, or approximately 40% of the data. Reliability was measured using a point-to-point comparison of the scores calculated by two transcribers. Overall agreement was considered to be acceptable at 80% (agreement by consonant: C1 = 89%, C2 = 69%, C3 = 84%, C4 = 82%).

The primary analysis was a mixed-design ANOVA with order as a between-subject effect and block and input frequency as within-subject effects. Talker variability was not included in this analysis because, as described above in the Materials section, multiple talkers were not possible for input

frequency 1, and the factor did not fully cross with input frequency. If production practice prevents perceptual learning, no input frequency effect should be observed in the test block of the baseline-test order. If production practice does not prevent perceptual learning, however, then the input frequency effect should be present during the test block in both the baseline-test and test-baseline orders.

If production practice has an effect, it should be observed in the order of blocks, specifically, greater accuracy in the second block compared to the first block. This effect was not expected to vary depending on the order condition.

Richtsmeier and Good did not observe an effect of talker variability, but benefits of this manipulation have been observed elsewhere (Plante, Bahl, Vance, & Gerken, 2011; Richtsmeier, Gerken, Goffman, & Hogan, 2009). Here, a mixed-design ANOVA was run using data from the test conditions for words with input frequencies of 3, 6, and 10, that is, the conditions in which talker variability was manipulated. In this analysis, talker variability was a within-subject effect and order was a between-subject effect. Order was included to account for the possibility that talker variability, like input frequency, might have no effect in the baseline-test condition.

These planned analyses were supplemented by post hoc analyses at the level of items. The perceptual learning and production practice effects were operationalized as difference scores and were calculated for each of the eight target items, that is, one learning effect per item, with separate calculations for the different order conditions, for a total of 16 data points. As an example calculation of a production practice effect for the baseline-test condition, average accuracy for /tʌvtʃəp/ at baseline was 9.94 and at test was 10.67, resulting in a production practice effect of 1.05. The scores were then entered into regression analyses with a sum of the ages of acquisition for each item's four consonants using acquisition data from McLeod and Crowe (2018; see also Moore, 2018; Moore, Fiez, & Tompkins, 2017 for experimental consonant age of acquisition effects in linguistic performance), as well as the

sums of phone and biphone probabilities calculated by the online phonotactic probability calculator (Vitevitch & Luce, 2004). Examples of these by-item phonological properties are presented in Table 3. Since /tʌvtʃəp/ had the highest consonant age of acquisition score at 13 (i.e., sum of consonant acquisition ages in years: /t/ = 3; /v/ = 4; /tʃ/ = 4; /p/ = 2), the baseline-test data point for /tʌvtʃəp/ can be seen in the top right corner of the left panel of Figure 3. No predictions were made for the post hoc analyses. However, as previously stated the analyses were motivated in part by the findings from Kaunshanskaya and Yoo (2011) in which production practice and perceptual learning effects varied depending on the phonological composition of the words. Phonotactic probability and consonant age of acquisition variables were selected for these post hoc analyses because they consistently have been shown in the literature to have a significant phoneme-level influence on linguistic tasks with perceptual and production components such as nonword repetition, among others (Moore et al., 2017; Munson, Edwards, & Beckman, 2011).

Although it was not considered as part of the primary analysis, children completed a referent identification task between the first block and second block. Children completed four trials, each featuring a horizontal array of three make-believe animals. The target and distracter animals were always drawn from the input frequency 3, 6, and 10 conditions. Children were asked to point to the animal corresponding to the sound file that was played at the beginning of the trial, and children received a point for each correctly identified animal. Accuracy means, standard deviations, 95% confidence intervals, and the results of a statistical comparison of the baseline-test and test-baseline conditions are reported in Table 2. We note that children who completed the test-baseline order had significantly more exposure to the test items when they completed the referent identification task. It is therefore sensible that they were significantly more accurate than children in the baseline-test order when completing this task.

Table 3 – Examples of how consonant age of acquisition sums, phone probability sums, and biphone probability sums were calculated for the nonwords /fugdən/ and /tʌvtʃəp/. Note that the phonotactic probability calculator automatically adds 1.0 to both the phone and biphone sums.

Nonwords	f	u	g	d	ə	n	Sums	t	ʌ	v	tʃ	ə	p	Sums
Consonant AoA^a	2		2	2		2	8	3		4	4		2	13
Phone Probabilities	0.0466	0.0221	0.0179	0.0403	0.0816	0.0832	1.2917	0.0445	0.0392	0.0236	0.0101	0.0816	0.0143	1.2133
Biphone Probabilities^b		0.0003	0.0002	0.0001	0.0022	0.0169	1.0197		0.0014	0.0012	0	0.0003	0.0022	1.0051

^aAoA = Age of acquisition. Ages are specific to consonants, so no ages are reported for vowels. ^bBiphone sums are reported for the combination of the column's sound and the previous sound. For example, the probability 0.0003 below the /u/ in /fugdən/ is the biphone probability for /fu/.

Results

Figure 2 presents a bar plot related to the main analysis, a $2 \times 2 \times 4$ mixed-design ANOVA (Order [baseline-test, test-baseline] \times Block [First Block, Second Block] \times Input Frequency [1, 3, 6, 10]).

Sphericity violations were observed for the main effect of input frequency and the interaction of input frequency and block, so a Greenhouse-Geisser correction was applied. There was no main effect of order, $p = .242$, but there was a main effect of block $F(1, 63) = 26.97, p < .001, \eta_p^2 = 0.30$ as participants were significantly more accurate when making productions in the second block ($M = 10.78$) compared to the first block ($M = 10.37$). There was a trend towards a main effect of input frequency, $F(2.49, 157.04) = 2.67, p = .060$, but more importantly, a significant three-way interaction of input frequency with block and order, $F(2.64, 165.98) = 6.53, p = .001, \eta_p^2 = 0.09$. We note briefly that, when children who repeated the exposures were removed from the analyses, the main effects and interaction reported above were again observed.

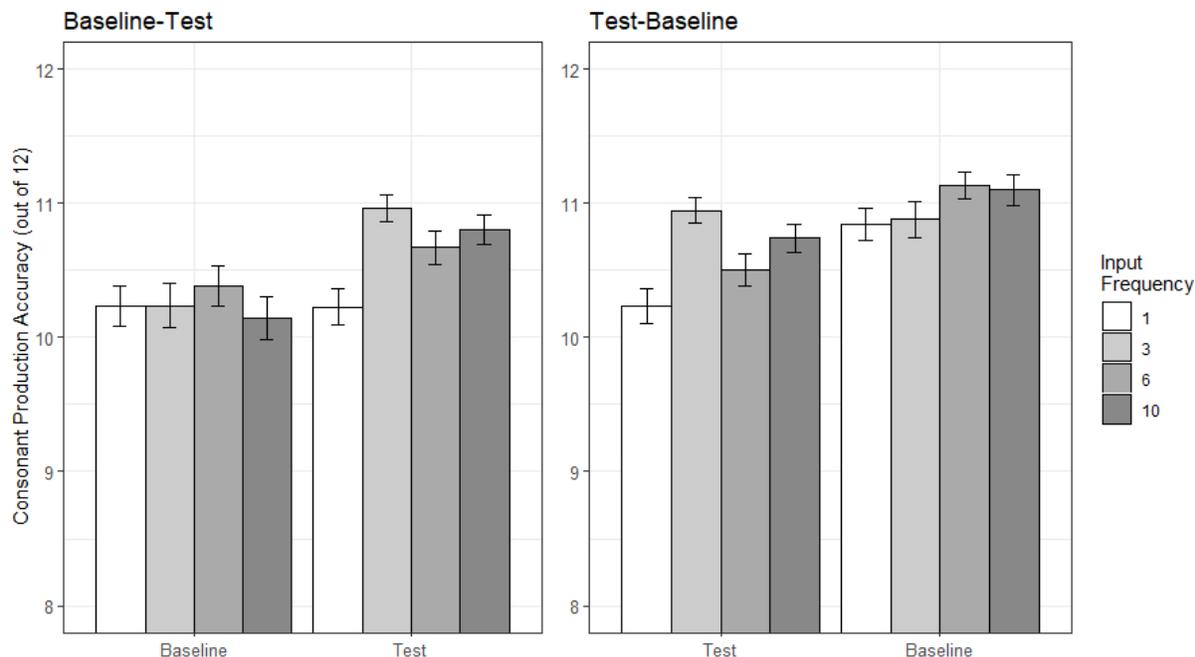


Figure 2 – Mean word accuracy (y-axis) broken down by block (x-axis), input frequency (shading of bars), and order (panels). Error bars reflect mean standard errors.

Figure 2 suggests that the interaction derives from an effect of input frequency that is observable only in the test blocks and regardless of order. To confirm, test data and baseline data were entered into two separate 2×4 mixed-design ANOVAs (Order \times Input Frequency). A Greenhouse-Geisser correction for sphericity was again applied. For the test data, there was no effect of order, $p = .632$, but there was a significant effect of input frequency, $F(2.70, 169.83) = 8.90$, $p < .001$, $\eta_p^2 = 0.12$. On average participants were most accurate during the test for items they heard 3 times ($M = 10.91$) followed by items they heard 10 times ($M = 10.73$), items they heard 6 times ($M = 10.54$), and finally items they heard 1 time ($M = 10.19$). There was no interaction between order and input frequency, $p = .854$. For the baseline data, there was a significant effect of order, $F(1, 63) = 6.88$, $p = .011$, $\eta_p^2 = 0.10$. Participants were more accurate when the baseline came after the test ($M = 10.90$) than when it came first ($M = 10.21$). There was no effect of input frequency, $p = .614$, and no interaction, $p = .484$. In sum, the three-

way interaction of input frequency, block, and order derived from an input frequency effect present only during test and regardless of whether the test was preceded or followed by a baseline.

To better understand the main effect of block (production practice) and the effect of input frequency during test (perceptual learning), two post hoc, by-item regression analyses were conducted to determine whether each effect was sensitive to phonological properties of the items. For the regression analysis of production practice (where the production practice effect was the difference between the first block and second block; model $R^2 = .741$, $SE = .19$), consonant age of acquisition was a significant predictor, $\beta = .14$, $t = 3.04$, $p = .010$, but not phone sums, $p = .513$, or biphone sums, $p = .698$. In general, items composed of later-acquired sounds benefitted more from production practice. For the analysis of perceptual learning (where the perceptual learning effect was the difference between input frequencies 1 and 3 in the test condition; model $R^2 = .494$, $SE = .44$), phone sums trended towards significance, $\beta = 5.93$, $t = 1.91$, $p = .081$, but not biphone sums, $p = .240$, or consonant age of acquisition, $p = .795$. Although the effect is only a trend and should be interpreted with caution, it is possible that items composed of frequent sounds benefitted more from perceptual learning. Scatter plots of the significant and near-significant correlations appear in Figure 3. The combined results of these regression analyses suggest that production practice and perceptual learning may be sensitive to different phonological aspects of novel words.

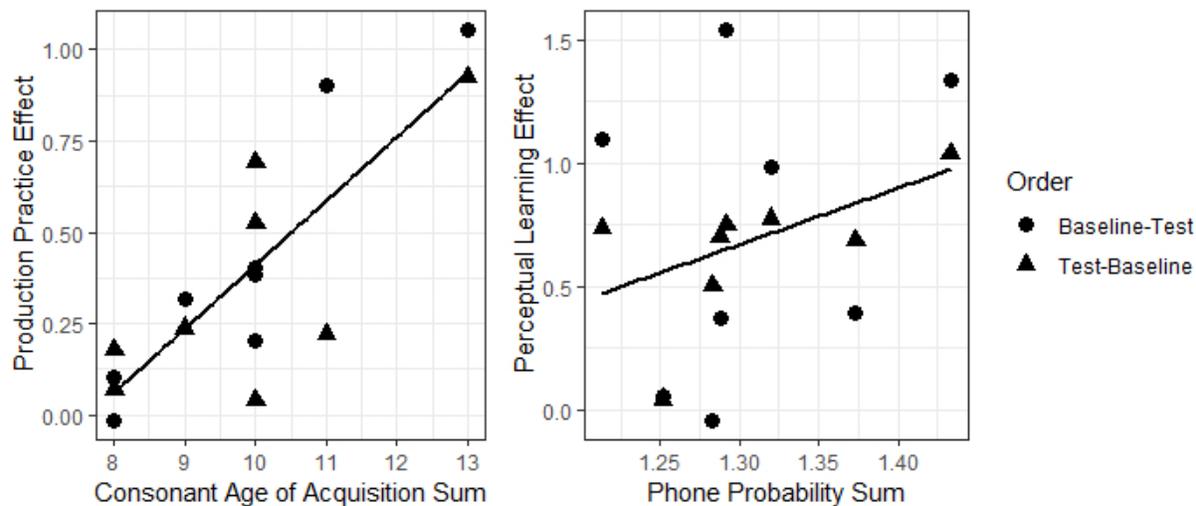


Figure 3 – Dot plots and regression lines for the correlations between (left panel) consonant age of acquisition sums and the production practice learning effect; (right panel) phone probability sums and the perceptual learning effect.

To determine whether talker variability affected production accuracy during the test phase, a 2×2 mixed-design ANOVA was conducted (Order [baseline-test, test-baseline], Talker Variability [single talker, multiple talkers]). There was no effect of order and no interaction with talker variability ($ps > .3$), but there was a main effect of talker variability, $F(1, 63) = 5.51, p = .022, \eta_p^2 = 0.08$. The results reveal a small benefit of hearing a novel word produced by multiple talkers ($M = 10.76$) compared to a single talker ($M = 10.57$). The direction of this effect is consistent with results reported by Richtsmeier and Good, but with 32 additional subjects, the observed power of the analysis improved from 0.25 to 0.65. The relatively small size of the effect is consistent with other work where talker variability proved to have limited benefits (Sinkeviciute, Brown, Brekelmans, & Wonnacott, 2019).

General Discussion

In this study, perceptual learning and production practice were examined as contributors to speech accuracy. The results reaffirm previous findings that both perceptual learning and production practice result in learning effects. Production practice was visible in the main effect of block: Participants were consistently more accurate in the second block. Perceptual learning was observed in the effect of input frequency during test: Participants were most accurate to produce items they first heard 3 times, and hearing an item 3, 6, or 10 times was superior to hearing it just once (see p. 2864 of the General Discussion of Richtsmeier & Good, 2018, for further consideration of the four levels of input frequency).

The specific question addressed in this manuscript was whether initial production practice negatively affects the benefits of perceptual learning to speech production accuracy. We predicted that—if production practice outweighs perceptual learning—no effect of input frequency should be found at test in the baseline-test order. Input frequency was significant in that condition, however, leaving two possible explanations for Richtsmeier and Good's finding that input frequency effects did not carry over to a posttest baseline. It may be that perceptual learning is simply a form of priming and limited in duration. However, it is also possible that perceptual learning is limited *and* production practice outweighs production practice over a larger set of productions. This latter conclusion is consistent with results reported by Richtsmeier and Goffman (2017), who found that children were more accurate to produce novel words supported by a high variability perceptual input, but only for the first 3 of 18 productions that children made. In sum, perceptual learning does appear to be a form of temporary priming, whereas production practice results in more consistent gains in speech accuracy. Future studies are needed to improve predictions for when these learning mechanisms will interact and to what degree one can interfere with the other.

Although we interpreted the higher accuracy in the second block to reflect a benefit of production practice, learners do not always benefit from speech production. Kaushanskaya and Yoo (2011) and Zamuner et al. (2018) both argue that production practice inhibits word learning when the words are composed of unfamiliar phonemes or phoneme sequences. However, the detriments observed by Kaushanskaya and Yoo and Zamuner et al. are specific to learning form-referent pairs, whereas the benefits of production practice observed in this study are specific to speech accuracy. It therefore appears that production practice can be both beneficial and detrimental, depending on the type of task employed. Given that there are many other aspects of language acquisition that production practice might affect, aspects such as intonation or sentence formulation, much remains to be understood about production's role in language acquisition.

The regression analysis of consonant age of acquisition also suggests additional nuance for our understanding of production practice. The regression analyses were not guided by predictions and should thus be treated cautiously. Nevertheless, the observed correlation between production practice and consonant age of acquisition suggests that production practice is most beneficial for sounds that are acquired later. This is perhaps surprising because novel words comprising later developing consonants often correspond with decreased performance in tasks such as nonword repetition, nonword reading, and lexical decision compared to words comprising earlier developing consonants (Moore, 2018; Moore et al., 2017; Moore, Tompkins, & Dollaghan, 2010). Moore and colleagues posit that consonant acquisition impacts how phonological information is stored in long-term memory. Nonwords containing later-acquired consonants appear to result in weaker activation of phonological information from long-term memory, are more prone to decay, and are therefore harder to maintain in short-term memory (Baddeley, 1986; Bowey, 2006; Edwards & Lahey, 1998), thus impacting language processing. We speculate that production practice may help to strengthen the weaker activations in phonological short-term memory so that items with later-acquired consonants show a greater learning benefit compared to

items with earlier-acquired consonants that presumably already have strong long-term-memory activations without production practice. We plan to test this hypothesis in future work.

We note two weaknesses of the study related to the participants. First, 14 children did not follow instructions and repeated the target items during the perceptual exposures. Although we would argue that this is a natural response to learning new words, it does make our analyses of perceptual learning and production practice more tenuous. Second, there were a number of unanticipated group differences between the participants in the test-baseline and baseline-test conditions, including greater age, lower auditory discrimination scores, lower nonword repetition scores, and marginally lower socioeconomic status in the baseline-test participants (see Table 2). One possible explanation for these differences is a difference in recruitment: Participants in the test-baseline condition were primarily recruited through newspaper ads and flyers handed out at daycares, whereas participants in the baseline-test condition were primarily recruited through social media. Regardless of the underlying causes for the group differences, we note that the critical finding here is a group similarity. That is, both groups benefitted from the input frequency manipulation during the test. Although it is possible that the results could have been different if the groups were more evenly matched, the consistency of the production practice and perceptual learning effects is striking.

Conclusion

The findings from this study support the claim that both perceptual learning and production practice facilitate word learning. Preschoolers' speech accuracy seemed to be primed briefly by perceptual learning, yet consistent gains in speech accuracy were observed in the analysis of production practice. However, the benefits of perceptual learning and production practice are not straightforward, and there is still a great deal to be learned about how these two learning processes play out during the larger window of speech development. Furthermore, these learning processes may depend on different phonological aspects of the words to be learned. Here, items that were more phonologically difficult

(based on age of acquisition of the consonants) showed a greater learning advantage from production practice compared to items that were less difficult. Future work could explore the extent of the interaction between these two learning mechanisms and how they may be involved differently across novel words with various lexical and phonological features. Conducting studies of this nature is critical for understanding how to provide maximally beneficial learning opportunities to children with speech and language delays.

Acknowledgments

The authors thank Paige Burgess, Jessica Filson, Amanda Good, and Bailey Rich for assistance recruiting and running participants.

References

- Baddeley, A. D. (1986). *Working Memory*. Oxford, UK: Clarendon Press.
- 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.
doi:<https://doi.org/10.1016/j.jml.2015.10.008>
- Bowey, J. A. (2006). Clarifying the phonological processing account of nonword repetition. *Applied Psycholinguistics*, *27*(4), 548-552. doi:<https://doi.org/10.1017/S0142716406220393>
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136-1146.
doi:<https://doi.org/10.1044/jslhr.4105.1136>
- 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.
doi:10.1044/1092-4388(2004/034)
- Edwards, J., & Lahey, M. (1998). Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics*, *19*(2), 279-309. doi:<https://doi.org/10.1017/S0142716400010079>
- Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation-2*. San Antonio, TX: Pearson.
- Kaushanskaya, M., & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, *26*(1), 121-148. doi:<https://doi.org/10.1080/01690965.2010.486579>
- McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology*, *27*(4), 1546-1571.
doi:https://doi.org/10.1044/2018_AJSLP-17-0100

- Moore, M. W. (2018). Consonant age of acquisition effects are robust in children's nonword repetition performance. *Applied Psycholinguistics*, 39(5), 933-959.
doi:<https://doi.org/10.1017/S014271641800005X>
- Moore, M. W., Fiez, J. A., & Tompkins, C. A. (2017). Consonant age-of-acquisition effects in nonword repetition are not articulatory in nature. *Journal of Speech, Language, and Hearing Research*, 60(11), 3198-3212. doi:https://doi.org/10.1044/2017_JSLHR-L-16-0359
- Moore, M. W., Tompkins, C. A., & Dollaghan, C. A. (2010). Manipulating articulatory demands in nonword repetition: a 'late-8' non-word repetition task. *Clinical Linguistics & Phonetics*, 24(12), 997-1008. doi:<https://doi.org/10.3109/02699206.2010.510917>
- Munson, B., Edwards, J., & Beckman, M. E. (2011). Phonological representations in language acquisition: Climbing the ladder of abstraction. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *Oxford Handbook of Laboratory Phonology* (pp. 288-309). Oxford, UK: Oxford University Press.
- Ohala, D. K. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders*, 32(6), 397-421; quiz 421-392. doi:[https://doi.org/10.1016/S0021-9924\(99\)00018-0](https://doi.org/10.1016/S0021-9924(99)00018-0)
- Paradigm. (2015). Paradigm Experimental Software (Version 2.4) [Computer software]. Lawrence, KS: Perception Research Systems Incorporated. Retrieved from <http://www.paradigmexperiments.com/>
- 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. doi:10.1016/j.jcomdis.2010.07.005
- Pritchard, V. E., Heron-Delaney, M., Malone, S. A., & MacLeod, C. M. (2019). The Production Effect Improves Memory in 7-to 10-Year-Old Children. *Child Development*.
doi:<https://doi.org/10.1111/cdev.13247>

Puranik, C. S., Schreiber, S., Estabrook, E., & O'Donnell, E. (2014). Comparison of name-writing rubrics: Is there a gold standard? *Assessment for Effective Intervention*, 40(1), 16-23.

doi:<https://doi.org/10.1177%2F1534508413502390>

Richtsmeier, P. T., Gerken, L., Goffman, L., & Hogan, T. (2009). Statistical frequency in perception affects children's lexical production. *Cognition*, 111(3), 372-377. doi:10.1016/j.cognition.2009.02.009

Richtsmeier, P. T., & Goffman, L. (2017). Perceptual statistical learning over one week in child speech production. *Journal of Communication Disorders*, 68, 70-80. doi:10.1016/j.jcomdis.2017.06.004

Richtsmeier, P. T., & Good, A. K. (2018). Frequencies in perception and production differentially affect child speech. *Journal of Speech, Language, and Hearing Research*, 61(12), 2854-2868.

doi:https://doi.org/10.1044/2018_JSLHR-S-17-0391

Sinkeviciute, R., Brown, H., Brekelmans, G., & Wonnacott, E. (2019). The role of input variability and learner age in second language vocabulary learning. *Studies in Second Language Acquisition*,

41(4), 795-820. doi:<https://doi.org/10.1017/S0272263119000263>

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36(3),

481-487. doi:<https://doi.org/10.3758/BF03195594>

Zamuner, T. S., Strahm, S., Morin-Lessard, E., & Page, M. P. (2018). Reverse production effect: Children recognize novel words better when they are heard rather than produced during training.

Developmental Science, 21(4), e126366. doi:<https://doi.org/10.1111/desc.12636>