Research Article

Speech Intervention Outcomes Associated With Word Lexicality and Intervention Intensity

Alycia Cummings,^a Janet Hallgrimson,^b and Sarah Robinson^c

Purpose: This study examined how lexical representations and intervention intensity affect phonological acquisition and generalization in children with speech sound disorders.

Method: Using a single-subject multiple baseline design, 24 children with speech sound disorders (3;6 to 6;10 [years; months]) were split into 3 word lexicality types targeting word-initial complex singleton phonemes: /ɹ ɪ tʃ θ/. Specifically, academic vocabulary words, nonwords (NWs), and high-frequency (HF) words were contrasted. Intervention intensity was examined by comparing the performance of 12 children who completed eleven 50-min sessions (4 children/word type) to the performance of 12 who completed 19 sessions (4 children/word type). Children's production accuracy of their treated phonemes and overall percent consonants correct values were used to measure phonological generalization via percentage accuracy scores and *d* scores.

Results: All word lexicality conditions elicited phonological change, suggesting that academic vocabulary words, NWs, and HF words are viable intervention targets. Group mean averages were similarly high for the NWs and HF words, although children in the NW condition demonstrated more consistent phonological gains. Children who received 19 intervention sessions achieved 6 times more gains in treated sound accuracy than did children who received 11 sessions.

Conclusions: Word lexicality did not significantly influence children's intervention outcomes. More intensive intervention, as characterized by the number sessions, resulted in greater phonological change than did a shorter intervention program. Intervention intensity outcomes should be considered when establishing best practices for speech intervention scheduling.

Supplemental Material: <https://doi.org/10.23641/asha.7336055>

For decades, speech sound disorder (SSD) intervention approaches have focused on the target phoneme(s) and/or the method of service delivery when considering intervention efficiency and effectiveness (for a review, see Williams, McLeod, & McCauley, 2010). Alternatively, the actual words containing treated phonemes are not often considered to be a factor in phonological learning (Storkel, 2018). Recently, researchers have suggested that the words used in intervention likely play a central role in how children perceive, produce, and eventually acquire phonological knowledge (Gierut, 2016; Storkel, 2018). That is, the lexical characteristics of words can potentially positively or negatively impact speech intervention outcomes.

Word Lexicality in Intervention

The development of sounds and words in young children occurs simultaneously (Vihman, 2017). This suggests that the development of sounds can influence lexical representations just as the lexicon can influence phonological learning. Current models of phonological representation specifically address the phonological/lexical interaction by identifying two levels of representation (Storkel & Morrisette, 2002; Werker & Curtin, 2005). The lexical representation corresponds to a word as a whole unit (e.g., /dag/ for “dog”), whereas the phonological representation corresponds to the individual sounds or sound sequences that make up the word (e.g., /d/, /a/, /g/, /da/, /ag/).¹ The lexical representation may play a role in how words are perceived and produced, whereas the phonological representation may affect how a (spoken) word is encoded and processed. Although speech intervention has traditionally addressed the bottom-up phonological representations of single phonemes, the effect of lexical representations

^aDepartment of Communication Sciences and Disorders, Idaho State University–Meridian

^bDepartment of Speech-Language Pathology, Northern Health Region Hospital, The Pas, Manitoba, Canada

^cDepartment of Communication Sciences and Disorders, University of North Dakota, Grand Forks

Correspondence to Alycia Cummings: cummalyc@isu.edu

Editor-in-Chief: Shelley Gray

Editor: Ignatius Nip

Received February 7, 2018

Revision received May 1, 2018

Accepted June 29, 2018

https://doi.org/10.1044/2018_LSHSS-18-0026

¹Werker and Curtin's (2005) Processing Rich Information from Multidimensional Interactive Representations (PRIMIR) model refers to these as the word form plane and phonemic plane.

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

on phonological learning is an issue requiring clinical attention.

A variety of lexical characteristics are present in all real words. For example, researchers have examined how the age of acquisition of a word (early vs. late), word frequency (high vs. low), and word neighborhood density (dense vs. sparse) impact word learning, in general, and phonological learning, specifically (see reviews by Gierut, 2016; Storkel, 2018). The effect of word lexicality is one focus of this article, with lexicality referring to whether or not a word is a real word or a nonword (NW). To this end, words along a continuum of lexicality were tested. Words with known lexical representations (i.e., high-frequency [HF] words) were placed on one end of the continuum, whereas words with no lexical representations (i.e., NWs) were placed on the other end; words with developing lexical representations (i.e., academic vocabulary [AV] words) were placed in the middle of the continuum.

HF Words in Intervention

Frequency refers to how often a word occurs in a language. A word can be quite common in its occurrence in verbal and written communication, or it can be very infrequent. The frequency of word occurrence in language may lead to unique intervention outcomes. HF words are commonly occurring. Thus, children have many opportunities to hear and practice them outside of intervention in daily interactions and activities, which makes them relevant, functional, and salient stimuli (Gierut, Morrisette, & Ziemer, 2010). HF words may reduce cognitive processing demands because they are more accessible than low-frequency words (Ambridge, Kidd, Rowland, & Theakston, 2015). The reduced processing demands of HF words may also facilitate intervention outcomes. That is, the ease of processing of the lexical representation allows neural resources to be devoted to learning beyond the targeted sound in treated words (Storkel, 2018).

To explain further, the two representational models (Storkel & Morrisette, 2002) suggest the lexical characteristics of real words may also help with sound generalization. Establishing and accessing a lexical representation may affect how a phonological representation is formed. Indeed, HF words have induced phonological change in both treated and untreated sounds across manner classes, whereas low-frequency words have been found to elicit minimal amounts of phonological generalization, such as only affecting untreated sounds (Gierut & Morrisette, 2012b). Therefore, the words used in intervention may impact phonological generalization patterns. The familiarity of HF words with salient lexical representations has potential to create greater phonological change than words without lexical representations, such as NWs.

NWs in Intervention

If the goal of speech intervention is to establish phonological representations of sounds produced in error, it

could be that the optimal target is a word that contains no lexical information, such as a phonotactically possible NWs. Although it may seem counterintuitive to teach a child with a (speech) disorder nonsense words that have no use in his or her language, NWs have the potential to be more effective in intervention than real words. For example, a child may have a “frozen” phonological representation (Bryan & Howard, 1992) for a word, in which case the child’s production of a sound in the frozen word might not change with intervention, even if his ability to produce that same sound in other words does improve. These frozen words could unintentionally and negatively affect the outcomes of intervention.

To explain, NWs may enhance phonological learning because the sounds, words, and connections between them are all being learned at the same time (Storkel, 2018). Children have no prior exposure or experience with NWs, so negative practice is nonexistent. As such, intervention involving NWs avoids possible bias associated with life experience (Cummings & Barlow, 2011; Gierut et al., 2010). In other words, NWs may help reduce cognitive processing demands, which may allow for more automatic sound and word productions. It is possible that intervention using real words may lead to slower phonological learning due to a hypothesized process of needing to eliminate the incorrect or erroneous phonological–lexical connection prior to establishing a correct phonological–lexical representation. By essentially having a blank slate, NWs may facilitate faster learning of intervention targets and potentially greater phonological learning overall (Storkel, 2018).

Following the predictions of the two-representation model of spoken word processing, using NWs in intervention allows for bottom-up language processing: The individual phonemes of the word can be targeted without the influence of lexical input. Essentially, NWs allow a child to focus attention exclusively on articulatory routines, without competition from syntactic, semantic, or lexical information because there is no associated lexical information (van Riper, 1939). Consistent with this hypothesis, NWs have proven efficacious in facilitating systemwide phonological gains (Gierut et al., 2010) and establishing treated target phonemes (Cummings & Barlow, 2011). Thus, it could be argued that unfamiliar sound combinations, such as those found in NWs, are beneficial intervention targets. Because of their nonlexical status, NWs could allow children to focus solely on the phonological form of intervention words.

AV Words in Intervention

Although NWs are viable intervention target options, the functionality of real words is important for children with SSD because they are often at risk for academic difficulties in language, reading, spelling, and math (Pennington & Bishop, 2009). Moreover, children with SSD are considered an at-risk population for vocabulary deficits because of their inability to produce novel new words composed of sounds that are out of their phonetic

inventory (Schwartz & Leonard, 1982). It has been suggested that increasing a child's expressive vocabulary also leads to increases in phonological diversity (Girolametto, Pearce, & Weitzman, 1997). This implies that the simple act of teaching children new words in speech intervention could help establish phonological knowledge in children with SSD.

Thus, a feasible option could be to use unfamiliar real words.

AV words fit these criteria of being real, meaningful words that are essentially unknown to young children. These words frequently occur across a variety of academic texts and literature (Beck, McKeown, & Kucan, 2013). Although they commonly occur in written text, AV words are low frequency in discourse especially in the speech of and to young children; language users acquire AV words when they enter an academic setting. Similar to NWs, children should be able to focus on the phonological forms of AV words due to their lack of prior exposure to the words and their phonological forms. Although AV words are arguably similar to NWs in their phonological familiarity and structure, AV words also have lexical characteristics like HF words, which NWs do not. Thus, AV words are an approximate intermediary lexical option, having both the novel phonological form of an NW and associated lexical content that could be learned over time.

Intervention Intensity

Although the words used in speech intervention might impact outcomes, the intensity of intervention could also positively or negatively affect the amount of phonological learning that occurs. The term *intervention intensity* has been described as a way of defining both the quality and quantity of learning experiences within and across sessions (Baker, 2012a; Warren, Fey, & Yoder, 2007). It is more complex than merely quantifying the occurrence of sessions over time. Instead, intervention intensity involves numerous components so that the exact number of teaching opportunities or episodes during an intervention program can be calculated. It is often assumed that more hours of speech intervention is better (Schooling, Venediktov, & Leech, 2010), with 15–20 required to change a speech pattern or error to make a difference in speech sound production (Jacoby, Lee, Kummer, Levin, & Craghead, 2002). However, more is not necessarily always better (Baker, 2012b), and it is necessary to explicitly compare and contrast intervention intensity variables to identify optimal recommendations.

The combination of intervention duration,² dose,³ and dose frequency⁴ results in a numerical measure of cumulative intervention intensity (e.g., 100 trials $\times$ 2 times/

week $\times$ 12 weeks = 2,400 trials). Thus, cumulative intervention intensity defines the total number of teaching episodes during a child's intervention program and is an important value for research programs to report so that other researchers and practicing speech-language pathologists (SLPs) can attempt to replicate the results. However, in a review of the current literature, only 11 of 146 studies provided sufficient data to calculate the cumulative intervention intensity value (Baker & Williams, 2011). In order to determine whether intervention intensity significantly impacts outcomes, intensity variables must be explicitly compared.

Research Questions

The first research question addressed by this study was “Do the lexical representations of words influence phonological change during speech intervention?” To address this question, children with SSD were enrolled in a speech intervention program that included HF words, AV words, or NWs presented in a storybook context. It was hypothesized that when presented with novel or non-sense sound [word] strings, children would engage solely in the processing of sounds without also processing the lexical representation. If this were the case, NWs and/or AV words might induce more phonological change than would HF words. Alternatively, in the absence of a target-appropriate phonological representation, lexical processing could support sound learning. Thus, the AV and/or HF words could facilitate greater phonological learning than would NWs.

The second research question addressed by this study was “Does intervention program duration affect phonological change?” To address this question, children were enrolled in either an 11- or 19-session intervention program. Based on the limited intervention intensity literature, it was predicted that children receiving more sessions would demonstrate greater systemwide phonological change than would children receiving fewer sessions.

Method

Participants

Twenty-four children (10 girls, 14 boys; ages 3;6 to 6;10 [years;months]) with SSD were recruited to participate in this speech intervention program. Recruitment fliers were posted in local preschools, child care centers, and monthly child activity advertisements e-mailed to school district parents. In addition, school-based SLPs were asked to refer children who had not yet begun receiving intervention for their SSD. Eight children were randomly assigned to each of three word lexicality conditions: AV (Children 1–8), NW (Children 9–16), and HF (Children 17–24). Within each word condition, four children completed 11 intervention sessions and four children completed 19 intervention sessions. Children were assigned to the 11- or 19-session programs based on when in the academic year their parents responded to the recruitment materials. Children whose parents responded between March and May were assigned

²Total intervention duration is the interval time for which an intervention is provided (e.g., 19 sessions or 12 weeks).

³Dose refers to the number of properly administered teaching episodes per intervention session (e.g., 100 trials per session).

⁴Dose frequency indicates the number of intervention sessions provided per time unit (e.g., twice a week).

to the 11-session summer program. Children whose parents responded between August and October were assigned to the 19-session academic year program. All children met the following entry criteria:

- no prior speech intervention;
- residency in a monolingual English-speaking household;
- a percentile score below 11 on the Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000);
- hearing within typical limits as determined by a standard audiometric screening (American National Standards Institute, 1991);
- an oral-peripheral mechanism exam completed within typical limits (Robbins & Klee, 1987);
- nonverbal cognitive skills meeting or exceeding the typical range (standard scores above 85) as assessed by the Brief IQ Screener on the Leiter International Performance Scale–Revised (Roid & Miller, 1997);
- receptive vocabulary skills meeting or exceeding the typical range (standard scores above 85) as assessed by the Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007);
- reduced consonant inventory, excluding a minimum of three English phonemes from their respective phonemic inventories (see Appendix). Specifically, phonemic status of a sound was established following the criterion of two unique sets of minimal pairs (e.g., “sing”–“ring” or “run”–“rub”), regardless of whether they were correct relative to adult production (Gierut, Simmerman, & Neumann, 1994).

Group averages of all standardized test scores were very similar across the three word lexicality conditions (see Table 1; individual participants’ scores are presented in the Appendix).

Experimental Design

A single-subject multiple baseline design, in which every child served as his or her own control, was used in this intervention program (McReynolds & Kearns, 1983). This design has been used effectively in many different types of intervention studies involving children with SSD (Cummings & Barlow, 2011; Gierut & Morrisette, 2015; Gierut et al., 2010). Following procedures for this design, the children with SSD were randomly assigned to one of three word lexicality groups: AV words, NWs, or HF words. Every child was evaluated in a baseline period in which no intervention was provided, followed by either 11 or 19 sessions of speech intervention. The 11-session intervention was a summer-based program, whereas the 19-session intervention occurred during the school year.

Each child first completed full assessments using the Assessment of English Phonology (AEP; Barlow, 2003) and the Little PEEP: Shorter Protocol for the Evaluation of English Phonotactics (PEEP; Barlow, 2012).⁵ Because of the length of both of these assessments, they were administered on separate days. Subsequent baselines consisted of shorter baseline AEP probes that specifically targeted each child’s target phoneme. Baseline measures were taken in an additional one to two sessions prior to the initiation of intervention. For the postintervention assessment, complete AEP and PEEP assessments were administered on separate days immediately after intervention was finished.

Stimuli

The stimulus presentation techniques used in this study were similar, regardless of the word lexicality condition. All word conditions used a storybook format for presenting treated words, which had the potential to establish some meaning for the word conditions. The AV words were taken from the high-incidence AV word list (Coxhead, 2000). To ensure that the children in the AV condition did not know their intervention words, prior to beginning the intervention, they were asked (a) if they had ever heard of each of the AV words, (b) if they could define the AV words, and (c) if they could use the AV word in a sentence or short story. Some of the children claimed that they had previously heard some of the words; however, none of the children could define or correctly use the words in sentences. Thus, the AV words were classified as being unknown (real) words (i.e., essentially NWs). The AV condition paired each word with a quick incidental learning (QUIL) definition (Oetting, Rice, & Swank, 1995). The QUIL definitions were designed to establish lexical representations for the AV words.

The NWs were created by altering the vowel and consonant phonetic characteristics of the AV words. NWs are always low in frequency, given their nonoccurrence in the language (by definition); thus, all NWs were unfamiliar to the children. No attempt to define the NWs was made, though they were presented within the story in the place of real words. It was assumed children were familiar with the lexical characteristics of the HF words, and they were presented within the storyline with no additional explanation. Thus, children were not asked to define NWs or HF words.

Two different word frequency databases were used to determine whether the word targets were high or low in frequency (Kučera & Francis, 1967): (a) written word database from the University of Washington, St. Louis (<http://128.252.27.56/Neighborhood/Home.asp>) and (b) the SUBTLEX_{US} online corpus (Brysbaert & New, 2009; <http://www.ugent.be/pp/experimentele-psychologie/en/research/documents/subtlexus/subtlexus3.zip>). The Kučera

⁵The AEP probe is available from the first author. The PEEP probe is freely available from <http://slpath.com/littlepeep.html>.

Table 1. Word type and intervention duration group averages of standardized test scores.

Condition	Age	GFTA2	GFTA%	KLPA2	PPVT4	Leiter-R
Academic vocabulary	4.91 (0.79)	63.13 (14.80)	4.25 (3.84)	60.88 (14.00)	102.13 (10.84)	118.88 (20.34)
Nonword	5.00 (1.37)	70.38 (10.28)	6.50 (3.02)	67.00 (17.32)	113.00 (7.07)	114.67 (9.54)
High frequency	4.96 (1.29)	61.00 (12.76)	4.13 (3.91)	62.00 (11.97)	100.43 (7.74)	103.86 (16.20)
11 Sessions	5.09 (1.08)	63.75 (12.79)	5.04 (4.13)	58.50 (12.06)	105.50 (10.54)	115.75 (16.73)
19 Sessions	4.76 (1.11)	67.00 (12.04)	4.88 (3.23)	66.42 (14.52)	106.50 (11.29)	108.08 (15.41)

Note. Standard deviation values are in parentheses. GFTA2 = standard score on the Goldman-Fristoe Test of Articulation–Second Edition; GFTA% = percentile score on the Goldman-Fristoe Test of Articulation–Second Edition; KLPA2 = standard score on the Kahn-Lewis Phonological Analysis–Second Edition (Khan & Lewis, 2002); PPVT4 = standard score on the Peabody Picture Vocabulary Test–Fourth Edition; Leiter-R = standard score on the Leiter International Performance Scale–Revised.

and Francis (1967) corpus has been frequently used in the child development literature; thus, using it was necessary to maintain consistency with previous studies. The SUBTLEX_{US} online corpus has a large corpus size and focuses on spoken words (based on television and film subtitles). Thus, the use of both corpora allows for the comparison of word frequency values from two commonly used databases.

For this study, a word with a frequency score of 100 or higher in the Kučera and Francis (1967) database was considered high frequency. Two HF words did not meet this criterion: thousand (97) and Thursday (33). Although Thursday had a relatively low-frequency count, it was included because many preschool-age and early elementary-age children discuss the days of the week on a regular basis. Four of the AV words had frequency scores over 100: labor (149), change (240), choice (113), and theory (129). In addition, a word with a log10 frequency score of 3.0 or higher in the SUBTLEX_{US} corpus was also considered high frequency; all HF words met this criterion. Although the vast majority of the AV words had log10 word frequency measures below 3.0, six AV words did qualify as being HF: change (4.089), choice (3.697), theory (3.164), route (3.037), challenge (3.006), and channel (3.096).

Other lexical and phonological variables were also considered and calculated. Age of acquisition, the period when a word enters a child's vocabulary (Sosa & Stoel-Gammon, 2012), was examined based on a database of over 30,000 English words (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012): <http://crr.ugent.be/archives/806>. Phonological neighborhood density and phonotactic probability (i.e., phoneme and biphone probabilities) were calculated using the CLEARPOND online corpus (Marian, Bartolotti, Chabal, & Shook, 2012): <http://clearpond.northwestern.edu/index.php>.

Factorial analyses of variance (ANOVAs) were completed to determine whether the phonological and lexical variables differed across word lexicality conditions (Supplemental Materials S1–S3). Word types did not differ by the number of phonemes ($p > .18$), phoneme probability ($p > .67$), and biphone probability ($p > .16$). The word types differed in terms of neighborhood density, $F(2, 57) = 4.908$, $p < .02$, $\eta^2 = .147$; post hoc pairwise comparisons

revealed that the HF words ($M = 8.45$) were significantly denser than the NWs ($M = 2.40$; $p < .009$). In addition, the HF words had higher log10 word frequency scores (3.99) than the AV words (2.78), $F(1, 38) = 41.460$, $p < .0001$. HF words also had higher (Kučera & Francis, 1967) frequency scores (244.65) than the AV words (55.55), $F(1, 38) = 18.617$, $p < .0001$. HF words were, on average, acquired earlier (5.47 years) than the AV words (8.24 years), $F(1, 38) = 24.087$, $p < .0001$. Thus, the AV and HF words varied significantly in terms of word frequency and age of acquisition, forming two very distinct lexical conditions.

Additional materials relevant to this article are located at the first author's website <http://www.bblab.org/wordtypes.php>. The website contains story text for the HF words, AV words, and NWs, as well as speech probe lists for the AEP (Barlow, 2003) and the PEEP (Barlow, 2012). Pictures created by the second author for the AV and NW stories and an example of an intervention session data tracking sheet are also provided. These additional materials are licensed under a Creative Commons Attribution–Noncommercial 4.0 International License. This license allows users to share and adapt our materials, on the condition that appropriate credit is given and that the use of the materials is not for commercial purposes. More information about the license is available at <https://creativecommons.org/licenses/by-nc/4.0/>.

Phoneme Targets

Every child's intervention program targeted a single phoneme that was excluded from his or her phonemic inventory. Each child was at most 10% stimutable (Miccio, Elbert, & Forrest, 1999) for his or her selected treated phoneme (see Table 2). The study's phoneme targets /ɪ, ɪ, ʃ, θ/ were selected based on the complexity theory of age acquisition and/or later-acquired sounds, which suggests that, by exposing a child to more complex sounds in intervention, other less complex sounds might also be acquired (Gierut, 2001, 2007). A licensed and certified SLP (either the first or second author) read all stories in person during the intervention sessions. Color picture cards made from illustrations in each story were introduced to the children in conjunction with the five treated words (Figure 1 provides examples

Table 2. Pre- and postintervention production accuracy measures and the percentage of accuracy change of the treated phoneme in untreated words on the speech probes.

Group	Number	Phoneme	Sessions	M_A	M_B	Accuracy change	SD_A	d	Word type group M_A	Word type group M_B	Word type group d
AV	1	l	11	1.50	4.90	3.40	2.42	3.71	1.63	11.03	10.28
	2	tʃ	11	1.96	1.96	0.00	3.39	0.00			
	3	ʒ	11	0.52	0.00	-0.52	0.61	-0.56			
	4	ʒ	19	5.90	33.86	27.96	3.34	30.58			
	5	ʒ	11	0.00	3.10	3.10	0.00	3.39			
	6	tʃ	19	0.00	24.85	24.85	0.00	27.18			
	7	l	19	2.27	6.68	4.41	4.54	4.82			
NW	8	ʒ	19	0.92	12.91	11.99	1.83	13.12	0.25	19.50	21.06
	9	ʒ	11	1.56	8.33	6.77	1.60	7.41			
	10	tʃ	19	0.00	3.92	3.92	0.00	4.29			
	11	ʒ	19	0.24	8.14	7.90	0.55	8.64			
	12	θ	11	0.00	33.33	33.33	0.00	36.46			
	13	l	11	0.00	6.79	6.79	0.00	7.43			
	14	tʃ	19	0.00	21.00	21.00	0.00	22.97			
HF	15	ʒ	11	0.20	0.48	0.28	0.45	0.31	0.23	20.19	21.83
	16	ʒ	19	0.00	74.05	74.05	0.00	81.00			
	17	θ	11	0.00	2.94	2.94	0.00	3.22			
	18	ʒ	11	0.00	0.00	0.00	0.00	0.00			
	19	tʃ	11	0.00	0.00	0.99	0.00	0.00			
	20	ʒ	11	0.00	0.00	0.00	0.00	0.00			
	21	l	19	0.00	0.81	0.81	0.00	0.89			
	22	tʃ	19	0.00	81.48	81.48	0.00	89.13			
	23	θ	19	1.85	67.52	65.67	3.21	71.83			
	24	θ	19	0.00	8.77	8.77	0.00	9.60			

Note. Effect size d measures for individual participants and word type groups are also presented. $SD_{A-pooled}$ across S_s value for all d calculations was 0.91. AV = academic vocabulary; NW = nonword; HF = high frequency.

of the /l/ words). These were used during intervention to elicit the children's productions.

Intervention Procedure

Intervention was delivered in two phases: Imitation and Spontaneous Production. This protocol was chosen in order to remain consistent with previous studies (Cummings & Barlow, 2011; Gierut & Morrisette, 2010; Morrisette & Gierut, 2002). Intervention was provided twice weekly for 50-min sessions, for 11 or 19 sessions over the course of 7–12 weeks.

Imitation Intervention Phase

Each Imitation session began with the reading of each child's assigned storybook. Participants in both groups received the same amount of exposure to the targeted words through their presented stories. The difference between word lexicality conditions was that the children in the AV group were also provided with a brief definition of each word through QUIL. Children in the NW and HF word conditions were presented their book without any additional teaching/learning experiences associated with the treated words.

Story time was followed by 10 min of direct placement and sound-shaping activities during which each child was given verbal, tactile, and physical cues to help elicit the phoneme (Secord, Boyce, Donohue, Fox, & Shine, 2007). The remainder of each session consisted of drill-play

activities of the child's choice. During the Imitation phase, each child repeated the clinician's verbal model until either achieving a preestablished performance-based or time-based criterion, whichever came first. Specifically, this phase continued until a child maintained 75% accurate production of the treated phoneme over two consecutive sessions (i.e., performance-based criterion) or until seven consecutive sessions were completed (i.e., time-based criterion).

Spontaneous Production Intervention Phase

During this phase of intervention, each child produced treated words without a model during drill-play activities. This phase continued until the child maintained either a performance-based criterion of 90% accuracy production of the treated phoneme over three consecutive sessions or a time-based criterion of four sessions (for the 11-session protocol) or 12 sessions (for the 19-session protocol).

Intervention Fidelity

All sessions were audio-recorded so that the authors could confirm that intervention for the children occurred in a similar manner. A research assistant reviewed one Imitation session (Supplemental Material S4) and one Spontaneous Production session (Supplemental Material S5) of each child to verify the following aspects of the study design: (a) read story targeting treated phoneme (Imitation

Figure 1. The academic vocabulary (AV) words, nonwords (NW), and high-frequency (HF) words used to target /ɹ/. The pictures used to elicit the AV words and NW were created by the second author. HF /ɹ/ words were elicited using pictures from the *Maisy* children's books (Cousins, 1999a, 1999b). Lists of the /ɹ ʃ θ/ words are presented in Supplemental Materials S1–S3.

Academic Vocabulary Words		Nonwords	High Frequency Words
“random”: /ɹændəm/ IL: “it means to not have a plan or reason”		“rendof”: /ɹendəf/	“reading”: /ɹɪdɪŋ/
“route”: /ɹaʊt/ IL: “it means a path, a way to get to a place”		“roudge”: /ɹaudʒ/	“ready”: /ɹedi/
“remove”: /ɹɪmʊv/ IL: “it means to take away”		“reloyze”: /ɹɪləʊzɪz/	“running”: /ɹʌnɪŋ/
“rigid”: /ɹɪdʒɪd/ IL: “it means stiff, not moving”		“rechit”: /ɹɪtʃɪt/	“really”: /ɹɪli/
“region”: /ɹɪdʒɪn/ IL: “it means part of a place”		“rechim”: /ɹɪtʃɪm/	“river”: /ɹɪvə/

phase) or a story related to the day's theme (Spontaneous Production phase), (b) shape treated phoneme for at least 10 min (Imitation phase only), (c) conduct three to six child-selected drill-play activities, and (d) elicit 100 productions of the treated words in each session. Although the first three aspects of the intervention fidelity protocols were consistently met in all children's sessions, the 100-trial target was not met in 21 of the selected 48 sessions (Supplemental Material S6). Sessions with fewer numbers of trials often involved younger participants with attention, behavioral, and/or motivational issues.

Reliability

To ensure transcription reliability, an estimate of interrater agreement was obtained. Trained listeners used the International Phonetic Alphabet to narrowly transcribe all speech samples using the PHON computer transcription and data analysis program (<https://www.phon.ca/phontrac>; Rose & MacWhinney, 2014). Using PHON's blind transcriber function, 100% of each child's AEP and PEEP speech samples were reliability-checked by a second transcriber; these speech samples were compared for point-by-point consonant agreement that was required to reach an agreement threshold of 85% for each sample. Overall, the

transcriber reliability was 91% (range: 85%–96%) based on 62,785 consonants transcribed.

Data Analysis

Daily Intervention Data

For the Imitation phase of both the 11- and 19-session programs, a 3 (word type: AV, NW, HF) between-subjects × 7 (sessions) within-subject factorial ANOVA was calculated comparing the treated phoneme accuracy in the treated words for participants in the word type conditions across their seven imitation sessions. For the Spontaneous Production phase, a 3 (word type) between-subjects × 4 (sessions) within-subject factorial ANOVA was completed for the 11-session group and a 3 (word type) × 12 (sessions) within-subject factorial ANOVA was completed for the 19-session group.

Generalization

Generalization was computed as percent accuracy scores based on the AEP and PEEP performance of each child measured pre- and postintervention. These data were derived from the phonetic transcriptions of the AEP and PEEP. Based on these transcriptions, the data were organized for standard descriptive phonological analysis according

to target sound and word position (Dinnsen, 1984) in PHON. Two production accuracy values were calculated for each child: (a) the treated phoneme in untreated words and (b) all consonants in untreated words (i.e., a percent consonants correct [PCC] analysis). Each consonant was point-by-point identified as being correct or incorrect in relation to its target phoneme. From these values, generalization data from each child were examined from both descriptive and quantitative perspectives. Generalization is descriptively defined using the learning criterion level of 10% or greater accuracy change in treated phonemes and/or PCC scores (Gierut & Morrisette, 2012a). For each child and each condition, generalization was examined to see if the 10% criterion was met.

Quantitatively, separate 3 (word type: AV, NW, HF) $\times$ 2 (session: 11, 19) between-subjects factorial ANOVAs were calculated comparing the treated phoneme accuracy or PCC scores for participants who were assigned to the word type conditions in either the 11- or 19-session programs. Specifically, the amount of change from pre- to post-intervention in treated phoneme accuracy (Intervention Change) or PCC scores (PCC Change) was calculated separately for each child. The Intervention Change or PCC Change scores were the dependent measures in the factorial ANOVAs.

Effect Sizes

To establish the relative magnitude of the treated phoneme generalization gains, an effect size measurement was included as an additional quantitative analysis. That is, effect size values can be compared across intervention conditions to identify which is relatively most efficacious. Although there are a variety of effect size measurements, the standard mean difference (SMD) was chosen as a conservative estimate of intervention effects (Busk & Serlin, 1992).

The SMD involves the computation of two means: the mean of the baseline data (M_A) and the mean of the generalization data (M_B ; Busk & Serlin, 1992). The difference in the two means ($M_B - M_A$) creates the numerator of the equation. The standard deviation (SD) of the baseline data (SD_A) is the denominator of the equation, which, upon division, results in the effect size d . The one difficulty with this equation is that often children in speech intervention studies have 0% accuracy preintervention and no variability in performance across baseline measures (leading to an SD of 0). An alternative to this equation is to pool all of the baseline data across participants to create an SD across subjects ($SD_{A-pooled\ across\ Ss}$; Gierut, Morrisette, & Dickinson, 2015). This new SD value then reflects the actual baseline variability for all of the children in a study. Thus, in this study, the revised SMD effect size was used to establish the relative magnitude of generalization. SMD d scores were calculated for each child separately and for word lexicality and intervention intensity groups. Gierut et al. (2015) identified mean d values for small, medium, and large effect sizes as $d = 1.40$, 3.61 , and 10.12 , respectively.

Correlation Analyses

The targeted dose or the number of teaching episodes per session was 100 trials per session. However, this dosage was not met by many children (Supplemental Material S6) because of the individual children's attentional, behavioral, and/or motivational issues. Thus, there was a considerable amount of variation in intervention dose within and across conditions. To account for this variation, the relationships between intervention dose and treated phoneme change and between intervention dose and PCC change were examined in correlational analyses.

Results

Daily Intervention Data

Daily data were tracked to demonstrate that learning of the targeted phoneme was taking place (see Figure 2). None of the children achieved the 75% accuracy criterion for the Imitation phase of intervention; thus, every child completed all assigned Imitation sessions. The number of trials per session varied by intervention program and word lexicality (Supplemental Material S6). In the 11-session program, on average the AV words elicited 104 responses (range: 30–187), NWs elicited 106 responses (range: 43–183), and HF words elicited 101 responses (range: 0–187). In the 19-session program, AV words elicited 66 responses (range: 18–87),⁶ NWs elicited 96 responses (range: 37–177), and HF words elicited 107 responses (range: 31–178).

None of the children achieved 90% accuracy for three consecutive Spontaneous Production sessions; thus, they completed all of their assigned sessions. Again, the number of trials per session varied by intervention program and word lexicality. In the 11-session intervention program, the AV words elicited, on average, 121 responses (range: 78–187), NWs elicited 115 responses (range: 66–173), and HF words elicited 121 responses (range: 72–161). In the 19-session intervention program, AV words elicited 76 responses (range: 44–102), NWs elicited 114 responses (range: 67–223), and HF words elicited 123 responses (range: 49–187).

11-Session Intervention Condition

For the children in the AV condition, mean accuracy of the treated phoneme in the treated words during Imitation was 9.14% and increased to 33.25% during Spontaneous Production. In the NW condition, mean accuracy during Imitation was 3.11% and 22.75% during Spontaneous Production. Children in the HF condition produced the treated phoneme with 7.89% average accuracy during Imitation, which increased to 13.13% during Spontaneous Production. No differences in phonological learning

⁶The number of responses for the 19-session AV condition was lower than for all other conditions. This was primarily due to two children (Nos. 5 and 6) who were exceptionally unmotivated and demonstrated a variety of poor behaviors that limited the number of trials completed in a session.

Figure 2. Individual children's daily intervention data separated into the 11-session (left) and 19-session (right) intervention conditions, organized by word lexicality type. Tx = treatment.

across word types were observed for the Imitation ($p > .61$) or Spontaneous Production ($p > .51$) phase.

19-Session Intervention Condition

During the AV condition, the mean accuracies were 5.07% during Imitation and 16.21% during Spontaneous Production. Children in the NW condition produced the treated phoneme with 8.54% accuracy during Imitation and 18.75% accuracy during Spontaneous Production. The mean accuracy average for children in the HF group was 8.96% during Imitation, which increased to 54.54% during Spontaneous Production. No word lexicality differences were observed for the Imitation phase ($p > .86$). A main effect of word type was observed during the Spontaneous Production phase, $F(2, 9) = 4.833$, $p < .04$, $\eta^2 = .518$. However, Bonferroni-controlled post hoc pairwise comparisons did not identify any significant differences between word types. Thus, although the children in the HF condition demonstrated much higher production accuracy scores of their phonemes in treated words than did the children in the AV and NW conditions, this was not a significant group difference.

Generalization

When considering generalization patterns, three children in the AV condition (4, 6, 8), three children in the NW condition (12, 14, 16), and two children in the HF condition (22, 23) demonstrated significant (McReynolds & Kearns, 1983) treated phoneme production accuracy gains of greater than 10% (see Table 2). One of these children was in the 11-session program (12), whereas the other seven children were in the 19-session program.

At a group level, children in the NW and HF word conditions and the 19-session program demonstrated intervention gains of greater than 10% (see Table 3). The factorial ANOVA identified no effect of word lexicality ($p > .58$). Thus, although the NW and HF groups demonstrated similar levels of treated phoneme accuracy, which were higher than those of the AV group, the group differences were not significant. The number of intervention sessions children received significantly impacted treated phoneme accuracy gains, $F(2, 18) = 6.335$, $p < .03$, $\eta^2 = .260$. Children who received 19 sessions (27.74%) demonstrated

gains that were, on average, over six times that of children who received 11 sessions ($M = 4.68\%$). Thus, receiving approximately 8 hr of additional speech intervention significantly improved children's ability to produce their treated phonemes. No Word Type $\times$ Session interaction was observed ($p > .51$).

In terms of overall PCC gains, four children in the AV condition (3, 4, 5, 8), five children in the NW condition (9, 11, 12, 14, 16), and four children in the HF condition (17, 22, 23, 24) demonstrated PCC production accuracy gains of greater than 10% (see Table 4). Of these children, five were in the 11-session program (3, 5, 9, 12, 17), whereas the other eight children were in the 19-session program.

Similar to the treated phoneme analysis, children in the NW and HF word conditions and the 19-session program demonstrated averaged group PCC gains greater than 10% (see Table 5). The factorial ANOVA identified no effect of word type ($p > .91$) or intervention session number ($p > .10$); no Word Type $\times$ Session interaction was observed ($p > .64$). Thus, the type of word used in intervention did not significantly change how much system-wide phonological change was observed. Although the children in the 19-session program (12.52%) had greater gains in PCC scores, they were not significantly different from the 11-session program (7.57%).

Effect Sizes

When considering the treated phoneme effect sizes for children individually, the children in the AV group had d scores ranging from -0.56 to 30.58 , children in the NW condition had d scores ranging from 0.31 to 81.00 , and children in the HF group received d scores ranging from 0 to 89.13 (see Table 2). Thus, children varied greatly in the magnitude of their learning their treated phoneme.

At the group level, the NW ($d = 21.06$) and HF ($d = 21.83$) conditions had similar effect sizes, with the AV condition having the lowest d score of 10.28 . Interpreting these treated phoneme effect sizes based on the criteria presented by Gierut et al. (2015), all three word lexicality conditions elicited large intervention effects. That both the NW and HF conditions had d scores double that of the AV condition suggest that they had a greater

Table 3. Mean treated phoneme change in untreated words for three word lexicality types and two intervention durations.

Word type	Intervention sessions		Overall
	11	19	
Academic vocabulary	1.495 (11.221)	17.305 (11.221)	9.400 (7.934)
Nonword	11.795 (11.221)	26.714 (11.221)	19.254 (7.934)
High frequency	0.735 (11.221)	39.183 (11.221)	19.959 (7.934)
Overall	4.675 (6.478)	27.734 (6.478)	n/a

Note. Standard error of the mean values are based on the 3 (word type) $\times$ 2 (session) between-subjects factorial analysis of variance.

Table 4. Pre- and postintervention and the percentage change of percent consonants correct scores on the speech probes.

Group	Number	Phoneme	Sessions	M_A	M_B	Accuracy change
AV	1	l	11	36.45	43.12	6.67
	2	tʃ	11	42.20	41.23	-0.97
	3	ɹ	11	46.93	60.85	13.92
	4	ɹ	19	72.54	85.01	12.47
	5	ɹ	11	56.39	71.31	14.92
	6	tʃ	19	20.51	22.01	1.50
	7	l	19	47.98	55.90	7.92
	8	ɹ	19	62.98	79.63	16.65
NW	9	ɹ	11	72.35	84.06	11.71
	10	tʃ	19	24.06	28.96	4.90
	11	ɹ	19	72.64	84.41	11.77
	12	θ	11	56.71	71.01	14.30
	13	l	11	33.75	35.80	2.05
	14	tʃ	19	47.80	65.77	17.97
	15	ɹ	11	66.53	65.66	-0.87
	16	ɹ	19	75.03	96.11	21.03
HF	17	θ	11	49.92	68.36	18.44
	18	ɹ	11	56.20	60.76	4.56
	19	tʃ	11	39.84	39.36	-0.48
	20	ɹ	11	55.75	62.32	6.57
	21	l	19	23.60	30.72	7.12
	22	tʃ	19	37.83	60.75	22.92
	23	θ	19	71.93	87.82	15.89
	24	θ	19	57.47	67.96	10.49

Note. AV = academic vocabulary; NW = nonword; HF = high frequency.

impact on children's phonological learning than did AV words.

Similar to the generalization analyses, the children in the 19-session program had a d score of 30.48, which was nearly six times larger than that of the 11-session program (5.13). This meant that the 19-session condition resulted in a large effect size whereas the 11-session condition resulted in a medium effect size. Thus, more intervention sessions resulted in greater treated phoneme accuracy gains.

Correlation Analyses

A positive correlation was found between intervention dose and treated phoneme change, $r(24) = .594, p < .003$. A positive correlation between the number of intervention trials completed and the change in PCC scores was also

identified, $r(24) = .435, p < .04$. That is, children's production accuracy of the treated phoneme and of all consonants improved as more trials were attempted in intervention. Thus, greater phonological change occurred as cumulative intervention intensity increased.

Discussion

This study tested how lexical representations impact the learning of phonemes, as predicted by the two representation models of word processing (Storkel & Morrisette, 2002). To accomplish this, three types of word lexicality were compared: AV words, NWs, and HF words. All three word lexicality types elicited positive phonological change in the children. Although word type group differences were not significantly different, the NWs and HF words elicited

Table 5. Mean percent consonants correct change in untreated words for three word lexicality types and two intervention durations.

Word type	Intervention sessions		Overall
	11	19	
Academic vocabulary	8.637 (3.599)	9.636 (3.599)	9.136 (2.545)
Nonword	6.799 (3.599)	13.931 (3.599)	10.365 (2.545)
High frequency	7.272 (3.599)	13.995 (3.599)	10.633 (2.545)
Overall	7.569 (2.078)	12.520 (2.078)	n/a

Note. Standard error of the mean values are based on the 3 (word type) $\times$ 2 (session) between-subjects factorial analysis of variance.

similar amounts of phonological change in both the treated phoneme and overall PCC, with values that were often nearly twice those of the AV words. Thus, NWs and HF words appear to be efficient and effective speech intervention target options.

The number of intervention sessions children received was also examined to see whether the duration of intervention impacted children's phonological change. Children who were enrolled in the 19-session program demonstrated significantly larger gains in the production accuracy of their treated phoneme, as compared to children in the 11-session program. Thus, the greater amount of practice time children had with their treated phoneme improved their ability to produce that sound in untreated words.

Word Lexicality Outcomes

Consistent with previous findings (Cummings & Barlow, 2011), the NWs and HF words resulted in similar levels of accuracy gains. Specifically, both word types might reduce lexical processing demands, which might increase phonological learning (Storkel, 2018). That is, when lexical processing demands are lessened, more language processing capacity can be devoted to the accurate production of speech sounds, which could result in greater amounts of phonological change (Gierut, 2016). However, it is possible that NWs and HF words function in different ways to achieve similar phonological outcomes.

Without any lexical information associated with NWs, children can only make use of the words' phonological information, which hypothetically allows them to focus on correct phoneme production. Processing demands are lessened because of NW's lack of established lexical representations. Thus, NWs facilitate phonological learning because the word, the phoneme, and the connection between the two are being learned simultaneously (Storkel, 2018). Although familiar words could be misarticulated due to their frequent occurrence (and incorrect production) in children's speech, no prior production behaviors are associated with NWs.

Alternatively, HF words have well-established lexical representations. Because of their common occurrence in the language, HF words are more accessible and faster to retrieve from the lexicon (Ambridge et al., 2015). Moreover, the familiarity of HF words makes them easier to remember. Thus, although HF words have lexical representations, the impact of the representations on sound learning could be negligible because of their ease of processing. Thus, HF words' reduced lexical processing demands potentially allow more cognitive capacity to be allocated to speech sound production (Storkel, 2018).

On the other hand, the lexical information associated with the AV words in this study could have been a distractor to phoneme learning. Recall that all three types of words were introduced within the storybook context, but the AV word exposures also included the QUIL exposures defining each word. The new lexical information associated with the AV words might have placed a heavy processing

load on the children, which may have resulted in cognitive resources being allocated to word learning, rather than to phonological processing. Future studies incorporating AV words that are not accompanied by a QUIL definition will provide additional information as to how lexical information impacts phonological learning. It is hypothesized that undefined AV words could function similarly to NWs, which would provide clinicians with an academically relevant intervention alternative to NWs.

One additional thing to consider is that, although the NW and HF group means were similar, the basis of those values was different. In general, the outcomes of the HF condition were driven almost completely by two high achieving children (22 and 23), whereas all children in the NW condition contributed to group outcomes. That is, the extreme values of the children in the HF condition averaged out to a mean value that was similar to that of the NW condition in which lower, but more consistent, scores were found across children. Thus, the children in the NW condition performed more consistently as a group than did the children in the HF word group.

Intervention Intensity Outcomes

This study contrasted the outcomes of two intervention durations, 11 sessions and 19 sessions, thus providing information regarding intervention intensity. Children in the 19-session duration group demonstrated significantly higher production accuracy of their treated phoneme than did children in the shorter duration group. This outcome was not unexpected, as an additional 400 min of intervention (8 sessions $\times$ 50 min) should provide children with extra practice time for their treated phoneme. This outcome also fits well with previous findings that children need at least 15 hr of intervention to make change in their SSD (Jacoby et al., 2002).

Importantly, this study adds to the intervention intensity literature by including effect sizes when comparing the duration groups. In order to identify optimal intervention dosage, both the characteristics of the intervention and the size of the effect must be known (Zeng, Law, & Lindsay, 2012). In this study, the effect size of the 19-session condition was approximately 30, whereas the effect size of the 11-session condition was approximately 5. If the goal is to achieve a large effect size of approximately 10 (Gierut et al., 2015), the findings from this study suggest that an optimal intervention duration should be more than 11 sessions but could likely be less than 19 sessions.

It is possible that the relationship between intervention intensity and intervention outcomes is not strictly linear and that there could be a point of diminishing returns when considering how much time should be allocated to any one phoneme (Baker, 2012b). This nonlinear relationship could also suggest that children may need to first achieve a minimum dosage (of some yet to be determined value) prior to achieving larger intervention gains. This hypothesis is supported by the observation that the 19-session condition had an effect size nearly six times that of the

11-session condition. If there was a true linear relationship between intervention duration and production accuracy, the effect size of the 19-session group should have been less than twice the size of the 11-session group. It is possible that the children in the shorter duration group were just beginning to demonstrate phonological learning and change when intervention was withdrawn. That is, the children might have established a base level of phonological knowledge but had not yet had the opportunity to intensely practice in order to fully establish and strengthen their phonological representations. This is akin to the idea of a prepractice phase in motor learning where learning to produce the sound occurs prior to a practice phase of intervention (Maas et al., 2008).

Limitations

Although children in the 19-session condition demonstrated greater changes in PCC scores than did the children in the shorter intervention condition, these differences were not significant. This could indicate that the amount of practice children received in intervention did not affect systemwide phonological change. However, alternative explanations are also possible. Specifically, although the number of sessions (11 or 19) was consistent within treatment groups, the number of trials (i.e., intervention dose) varied greatly from session to session and child to child. The correlation analyses addressed this dosage variability and revealed that more phonological learning of the treated phoneme alone and of all consonants occurred with more intervention trial attempts. Thus, speech intervention of greater intensity appears to be more effective in eliciting phonological change than less intensive intervention.

Variability in phonological learning within and between groups was observed. One possible explanation for the variation lies with the fact that four different complex consonants were targeted in intervention. Because treated phonemes were chosen based on each child's stimulability levels and preintervention phonemic inventory, not every child was missing the same phoneme. Moreover, because the children were randomly assigned to each word lexicality condition, target phonemes in each group were not consistent: 10 /ɪ/, six /ʃ/, and four each of /l/ and /θ/. Breaking down the analyses into four treated phonemes, three word lexicality types, and two intervention durations would not have been informative with the current small sample. Future studies directly controlling these variables are needed to determine whether some phoneme, word, and/or intervention duration combinations lead to better outcomes than others.

The session duration of all sessions in this study was 50 min. Although this is a common duration for research programs, it is not a functional session duration for most school-based and private clinic speech intervention programs, where children are often seen for 15- to 30-min sessions. Intervention outcomes of shorter duration sessions are currently unknown. An initial hypothesis based on this study would be that more sessions would lead to greater

phonological change, regardless of the session duration. However, it is possible that the more important intensity factor is the number of trials attempted, rather than the number of sessions children attend. Future studies involving shorter intervention sessions should be completed. An additional intervention option would be to explicitly control the number of trials occurring in a session (Krueger, 2017). That is, children would remain in a session only for the amount of time required to complete a set number of trials. Thus, the session duration would vary, though the cumulative intervention intensity would remain the same.

Clinical Implications

Gierut (2016) suggested that certain combinations of word targets may be more efficacious than others. This study found that NWs and HF words were similarly effective targets in SSD intervention. However, it is possible that a combination of word lexicality types might result in greater outcomes than intervention involving just one word type. For example, NWs could establish the phonological representation of a phoneme, and then meaningful HF words could extend the learning to everyday situations. Alternatively, AV word intervention could follow NW intervention, which would continue to provide children with unfamiliar phonological representations. If the AV words were provided with definitions, this could also expand children's expressive vocabulary skills and phonological diversity (Girolametto et al., 1997). For example, pilot data suggest that children treated first with NWs and then with AV words demonstrated larger phonological gains than children treated with AV words followed by NWs (Cummings, 2017). More information regarding the combination of word lexicality types in intervention is necessary.

The results of this study provide initial guidelines regarding intervention scheduling of children with SSD. That is, to achieve a large, measurable intervention effect, children must complete more than eleven 50-min sessions. In addition, greater production accuracy was correlated with more trial attempts. Together, these observations suggest, in order for speech intervention to be effective, it must also be intensive in nature. Whether intervention intensity is best captured by the number of sessions allocated or the total number of trials attempted is yet to be discovered.

Acknowledgments

This research was supported by the National Institute on Deafness and Other Communication Disorders Grant R15DC013359 and the National Center for Research Resources Grant C06RR022088 awarded to the first author. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to thank Maria Fairfield, Steph Frey, Kristen Giesbrecht, Kristyne Hass, Brianna Jallo, Jessie Lancaster, Tara Lundmark, Abbi Machart, Anna Morris, Megan Nauman, Courtney Rowan, and Andrea Watkins for help with speech transcription. In addition,

Kristen Giesbrecht provided manuscript suggestions regarding intervention intensity. The authors also thank Amy Hardy and Mary van Donsel for reading previous drafts of the article.

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>
- American National Standards Institute. (1991). *American national standard specifications for audiometers* (ANSI S3.6-1969). New York, NY: Author.
- Baker, E. (2012a). Optimal intervention intensity. *International Journal of Speech-Language Pathology*, 14(5), 401–409. <https://doi.org/10.3109/17549507.2012.700323>
- Baker, E. (2012b). Optimal intervention intensity in speech-language pathology: Discoveries, challenges, and uncharted territories. *International Journal of Speech-Language Pathology*, 14(5), 478–485. <https://doi.org/10.3109/17549507.2012.717967>
- Baker, E., & Williams, A. L. (2011). *Intervention intensity for speech sound disorders: How much and for how long?* Paper presented at the American Speech-Language-Hearing Association Convention, San Diego, CA.
- Barlow, J. A. (2003). *The Assessment of English Phonology (AEP)*. San Diego, CA: San Diego State University.
- Barlow, J. A. (2012). *Little PEEP: Shorter Protocol for the Evaluation of English Phonotactics*. San Diego, CA: San Diego State University.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford Press.
- Bryan, A., & Howard, D. (1992). Frozen phonology thawed: The analysis and remediation of a developmental disorder of real word phonology. *International Journal of Language & Communication Disorders*, 27(4), 343–365. <https://doi.org/10.3109/13682829209012045>
- Bryshaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>
- Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis* (pp. 187–212). Hillsdale, NJ: Erlbaum.
- Cousins, L. (1999a). *Maisy goes shopping*. Somerville, MA: Candlewick.
- Cousins, L. (1999b). *Maisy's bedtime*. Somerville, MA: Candlewick.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>
- Cummings, A. (2017). *Longitudinal evidence supporting the effectiveness of complexity treatment for speech sound disorders*. Paper presented at the American Speech-Language-Hearing Association Convention, Los Angeles, CA.
- Cummings, A., & Barlow, J. A. (2011). A comparison of word lexicality in the treatment of speech sound disorders. *Clinical Linguistics & Phonetics*, 25(4), 265–286. <https://doi.org/10.3109/02699206.2010.528822>
- Dinnsen, D. A. (1984). Methods and empirical issues in analyzing functional misarticulation. *ASHA Monographs*, 22, 5–17.
- Dunn, L., & Dunn, D. (2007). *Peabody Picture Vocabulary Test—Fourth Edition*. Circle Pines, MN: AGS/Pearson Assessments.
- Gierut, J. A. (2001). Complexity in phonological treatment: Clinical factors. *Language, Speech, and Hearing Services in Schools*, 32(4), 229–241. [https://doi.org/10.1044/0161-1461\(2001/021\)](https://doi.org/10.1044/0161-1461(2001/021))
- Gierut, J. A. (2007). Phonological complexity and language learnability. *American Journal of Speech-Language Pathology*, 16(1), 6–17. [https://doi.org/10.1044/1058-0360\(2007/003\)](https://doi.org/10.1044/1058-0360(2007/003))
- Gierut, J. A. (2016). Nexus to Lexis: Phonological disorders in children. *Seminars in Speech and Language*, 37(4), 280–290. <https://doi.org/10.1055/s-0036-1587704>
- Gierut, J. A., & Morrisette, M. L. (2010). Phonological learning and lexicality of treated stimuli. *Clinical Linguistics & Phonetics*, 24(2), 122–140. <https://doi.org/10.3109/02699200903440975>
- Gierut, J. A., & Morrisette, M. L. (2012a). Age of word acquisition effects in treatment of children with phonological delays. *Applied Psycholinguistics*, 33(1), 121–144. <https://doi.org/10.1017/S0142716411000294>
- Gierut, J. A., & Morrisette, M. L. (2012b). Density, frequency and the expressive phonology of children with phonological delay. *Journal of Child Language*, 39(4), 804–834. <https://doi.org/10.1017/S0305000911000304>
- Gierut, J. A., & Morrisette, M. L. (2015). Dense neighborhoods and mechanisms of learning: Evidence from children with phonological delay. *Journal of Child Language*, 42(5), 1036–1072. <https://doi.org/10.1017/S0305000914000701>
- Gierut, J. A., Morrisette, M. L., & Dickinson, S. L. (2015). Effect size for single-subject design in phonological treatment. *Journal of Speech, Language, and Hearing Research*, 58(5), 1464–1481. https://doi.org/10.1044/2015_JSLHR-S-14-0299
- Gierut, J. A., Morrisette, M. L., & Ziemer, S. M. (2010). Non-words and generalization in children with phonological disorders. *American Journal of Speech-Language Pathology*, 19(2), 167–177. [https://doi.org/10.1044/1058-0360\(2009/09-0020\)](https://doi.org/10.1044/1058-0360(2009/09-0020))
- Gierut, J. A., Simmerman, C. L., & Neumann, H. J. (1994). Phonemic structures of delayed phonological systems. *Journal of Child Language*, 21(2), 291–316.
- Girolametto, L., Pearce, P. S., & Weitzman, E. (1997). Effects of lexical intervention on the phonology of late talkers. *Journal of Speech, Language, and Hearing Research*, 40(2), 338–348.
- Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation—Second Edition*. Minneapolis, MN: Pearson.
- Jacoby, G. P., Lee, L., Kummer, A. W., Levin, L., & Creaghead, N. A. (2002). The number of individual treatment units necessary to facilitate functional communication improvements in the speech and language of young children. *American Journal of Speech-Language Pathology*, 11(4), 370–380. [https://doi.org/10.1044/1058-0360\(2002/041\)](https://doi.org/10.1044/1058-0360(2002/041))
- Khan, L., & Lewis, N. (2002). *Khan-Lewis Phonological Analysis—Second Edition (KLPA2)*. Circle Pines, MN: American Guidance Service.
- Krueger, B. I. (2017). *When to start: The consideration of age in the selection of treatment targets*. Paper presented at the American Speech-Language-Hearing Association Convention, Los Angeles, CA.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuperman, V., Stadthagen-Gonzalez, H., & Bryshaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4), 978–990. <https://doi.org/10.3758/s13428-012-0210-4>
- 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\(2008/025\)](https://doi.org/10.1044/1058-0360(2008/025))
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A.** (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One*, 7(8), e43230. <https://doi.org/10.1371/journal.pone.0043230>
- McReynolds, L. V., & Kearns, K. P.** (1983). *Single-subject experimental designs in communicative disorders*. Austin, TX: Pro-Ed.
- Miccio, A. W., Elbert, M., & Forrest, K.** (1999). The relationship between stimulability and phonological acquisition in children with normally developing and disordered phonologies. *American Journal of Speech-Language Pathology*, 8, 347–363.
- Morrisette, M. L., & Gierut, J. A.** (2002). Lexical organization and phonological change in treatment. *Journal of Speech, Language, and Hearing Research*, 45(1), 143–159. [https://doi.org/10.1044/1092-4388\(2002/011\)](https://doi.org/10.1044/1092-4388(2002/011))
- Oetting, J. B., Rice, M. L., & Swank, L. K.** (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research*, 38(2), 434–445.
- Pennington, B. F., & Bishop, D. V. M.** (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, 60, 283–306. <https://doi.org/10.1146/annurev.psych.60.110707.163548>
- Robbins, J., & Klee, T.** (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders*, 52(3), 271–277.
- Roid, G., & Miller, L.** (1997). *Leiter International Performance Scale-Revised (Leiter-R)*. Wood Dale, IL: Stoelting.
- Rose, Y., & MacWhinney, B.** (2014). The PhonBank Project: Data and software-assisted methods for the study of phonology and phonological development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford Handbook of Corpus Phonology* (pp. 308–401). Oxford, United Kingdom: Oxford University Press.
- Schooling, T., Venediktov, R., & Leech, H.** (2010). *Evidence-based systematic review: Effects of service delivery on the speech and language skills of children from birth to 5 years of age*. Rockville, MD: American Speech-Language-Hearing Association.
- Schwartz, R. G., & Leonard, L. B.** (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, 9(2), 319–336. <https://doi.org/10.1017/S0305000900004748>
- Secord, W. A., Boyce, S. E., Donohue, J. S., Fox, R. A., & Shine, R. E.** (2007). *Eliciting sounds: Techniques and strategies for clinicians* (2nd ed.). Clifton Park, NY: Cengage Learning.
- Sosa, A. V., & Stoel-Gammon, C.** (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research*, 55(2), 596–608. [https://doi.org/10.1044/1092-4388\(2011/10-0113\)](https://doi.org/10.1044/1092-4388(2011/10-0113))
- Storkel, H. L.** (2018). Implementing evidence-based practice: Selecting treatment words to boost phonological learning. *Language, Speech, and Hearing Services in Schools*, 49(3), 482–496. https://doi.org/10.1044/2017_LSHSS-17-0080
- Storkel, H. L., & Morrisette, M. L.** (2002). The lexicon and phonology: Interactions in language acquisition. *Language, Speech, and Hearing Services in Schools*, 33(1), 24–37. [https://doi.org/10.1044/0161-1461\(2002/003\)](https://doi.org/10.1044/0161-1461(2002/003))
- van Riper, C.** (1939). *Speech correction principles and methods*. Englewood Cliffs, NJ: Prentice Hall.
- Vihman, M. M.** (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, 108(1), 1–27. <https://doi.org/10.1111/bjop.12207>
- Warren, S. F., Fey, M. E., & Yoder, P. J.** (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, 13(1), 70–77. <https://doi.org/10.1002/mrdd.20139>
- Werker, J. F., & Curtin, S.** (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1(2), 197–234. <https://doi.org/10.1080/15475441.2005.9684216>
- Williams, A. L., McLeod, S., & McCauley, R.** (2010). *Interventions for speech sound disorders in children*. Baltimore, MD: Brookes.
- Zeng, B., Law, J., & Lindsay, G.** (2012). Characterizing optimal intervention intensity: The relationship between dosage and effect size in interventions for children with developmental speech and language difficulties. *International Journal of Speech-Language Pathology*, 14(5), 471–477. <https://doi.org/10.3109/17549507.2012.720281>

Appendix

Individual Children's Standardized Test Scores

Number	Phoneme	Condition	Tx		Age	GFTA2	GFTA%	KLPA2	PPVT4	Leiter-R	Phonemes missing
			sessions	Age							
1	l ^a	Academic vocabulary	11	3.92	77.00	9.00	56.00	95.00	113.00	g θ ð s z ʃ ʒ ʧ ʤ l ɹ	
2	tʃ ^a		11	5.67	40.00	0.50	40.00	90.00	102.00	θ ð s z ʃ ʒ ʧ ʤ	
3	ɹ		11	5.42	66.00	2.00	69.00	99.00	117.00	θ ð z ʃ ʒ ʧ ʤ ɹ	
4	ɹ ^a		19	5.83	71.00	3.00	75.00	104.00	143.00	v z ʒ ʤ ɹ j h	
5	ɹ	Nonword	11	4.75	80.00	11.00	72.00	119.00	153.00	z ʒ ɹ	
6	tʃ		19	4.08	42.00	0.50	40.00	96.00	107.00	k g f v θ ð ʃ ʒ ɳ l ɹ	
7	l ^a		19	4.08	65.00	4.00	68.00	96.00	93.00	g z ʃ ʒ ʧ ʤ ɹ j	
8	ɹ		19	5.50	64.00	4.00	67.00	118.00	123.00	g ʃ ʒ ʤ ɹ	
9	ɹ ^a		11	6.83	65.00	5.00	73.00	117.00	129.00	g v s z ʒ m ɹ j	
10	tʃ		19	3.67	55.00	2.00	40.00	110.00	100.00	k g f v θ ð s z ʃ ʒ ʧ ʤ ɳ l ɹ h	
11	ɹ		19	5.42	86.00	8.00	88.00	112.00	115.00	g f v s z ʃ ʒ ʤ m j ɹ h	
12	θ		11	4.17	74.00	10.00	73.00	103.00	121.00	g θ ð z ʃ ʒ ɳ l ɹ	
13	l		11	3.50	59.00	3.00	40.00	117.00	105.00	k g v θ ð z ʃ ʒ ʧ l ɹ	
14	tʃ		19	3.92	70.00	6.00	78.00	121.00	105.00	g ð z ʃ ʒ ʧ ʤ ɳ ɹ h	
15	ɹ	High frequency	11	5.67	77.00	10.00	64.00	118.00	113.00	g z ʒ ɳ ɹ j	
16	ɹ		19	6.83	77.00	8.00	60.00	126.00	117.00	p v z ʃ ʒ ʧ ʤ m l ɹ j h	
17	θ		11	3.92	69.00	7.00	57.00	100.00	103.00	g v θ ð z ʒ ʧ ʤ ɳ ɹ	
18	ɹ		11	6.08	57.00	1.00	47.00	113.00	129.00	p g s z ʃ ʒ ʤ ɹ	
19	tʃ		11	4.75	56.00	1.00	58.00	96.00	87.00	t g θ ð z ʃ ʒ ʧ ʤ l ɹ	
20	ɹ		11	6.42	45.00	1.00	53.00	99.00	117.00	z ʃ ʒ ʧ ʤ ɹ	
21	l		19	3.42	55.00	1.00	62.00	91.00	111.00	k g v z ʒ ʤ ɳ l w	
22	tʃ		19	3.50	76.00	9.00	77.00	103.00	103.00	k g θ ð z ʃ ʒ ʧ ʤ ɳ l ɹ	
23	θ ^a		19	5.58	77.00	10.00	75.00	107.00	85.00	k g θ ð z ʒ ɳ ɹ	
24	θ		19	5.33	66.00	3.00	67.00	94.00	95.00	g θ ð ʒ l ɹ h	

Note. Tx = treatment; GFTA2 = standard score on the Goldman-Fristoe Test of Articulation–Second Edition; GFTA% = percentile score on the Goldman-Fristoe Test of Articulation–Second Edition; KLPA2 = standard score on the Kahn-Lewis Phonological Analysis–Second Edition; PPVT4 = standard score on the Peabody Picture Vocabulary Test–Fourth Edition; Leiter-R = standard score on the Leiter International Performance Scale–Revised.

^aChild was 10% stimulable (Miccio et al., 1999) for treated phoneme; all other children were 0% stimulable for treated phoneme.