

Age of word acquisition effects in treatment of children with phonological delays

JUDITH A. GIERUT and MICHELE L. MORRISSETTE
Indiana University

Received: December 3, 2008 Accepted for publication: May 2, 2010

ADDRESS FOR CORRESPONDENCE

Judith A. Gierut, Department of Speech and Hearing Sciences, 200 South Jordan Avenue,
Bloomington, IN 47405-7002. E-mail: gierut@indiana.edu

ABSTRACT

The effects of the age of acquisition (AoA) of words were examined in the clinical treatment of 10 preschool children with phonological delays. Using a single-subject multiple-baseline experimental design, children were enrolled in one of four conditions that varied the AoA of the treated words (early vs. late acquired) relative to their corresponding word frequency (high vs. low frequency). Phonological generalization to treated and untreated sounds in error served as the dependent variable. Results showed that late acquired words induced greater generalization, with an effect size four times greater than early acquired words, whereas the effects of word frequency were minimized. Results are discussed relative to hypotheses about the role of AoA in language acquisition and the relevance of this variable for phonological learning.

The study of phonological acquisition has benefited from new insights about the structure, organization, and development of the mental lexicon. An emerging view is that phonology and the lexicon interact, such that words as units are as influential to the process of phonological learning as are the individual sounds that make up those words (Storkel & Morrisette, 2002). This paper builds on this perspective by exploring how the estimated age of acquisition (AoA) of a word bears on phonological learning by preschool children who present with phonological delays. This population affords a unique window onto the process of phonological learning because the children have reduced consonantal inventories that contribute to unintelligibility and that warrant treatment. Treatment may be designed as an experiment to induce learning, allowing for a carefully controlled evaluation of the variables that contribute to production accuracy and accelerate generalization to treated and untreated sounds in error. It is in this context that we evaluate the impact of AoA.

The theoretical motivation for the work builds on models of spoken word recognition, which have isolated various attributes of words that influence behavioral

performance in perception, production, and lexical retrieval. Although the literature on this topic is burgeoning with respect to the performance of adults and typical children, few studies have explored how the attributes of words affect the ways in which children with phonological delays acquire the sound system in the context of treatment. To our knowledge, neighborhood density and word frequency are the only parameters that have been evaluated experimentally from a treatment vantage for children with phonological delays, and thus are the focus of the review.

NEIGHBORHOOD DENSITY AND PHONOLOGICAL TREATMENT

One working hypothesis is that words of the lexicon are organized into neighborhoods. Neighborhoods consist of phonetically similar sounding words that differ from each other by a one phoneme substitution, deletion, or addition, for example, “rat” is a neighbor of “bat,” “at,” and “frat,” among others (Luce & Pisoni, 1998). The number of neighbors to a given word varies, thereby yielding dense versus sparse pockets of the lexicon. This organization has consequences for spoken word recognition: words from dense neighborhoods have an inhibitory effect on performance in studies of adults, with slower and less accurate responding on a variety of psycholinguistic tasks (e.g., Luce, Pisoni, & Goldinger, 1990; Munson & Solomon, 2004; Vitevitch, 1997).

In typical development, however, infants and children appear to benefit in lexical development from dense neighborhood structure (Charles-Luce & Luce, 1990; Jusczyk, Luce, & Charles-Luce, 1994; Walley, Metsala, & Garlock, 2003). One thought is that children’s early words are stored with only the coarsest phonological structure (e.g., consonant–vowel–consonant). As new words are added to the lexicon, the density of lexical neighborhoods increases accordingly. Because of the increased density, children are forced to elaborate the phonological structure of words in order to uniquely differentiate one form from another. Increases in vocabulary size lead to increases in neighborhood density, which in turn prompt the reorganization and refinement of the phonological quality of lexical representations. This premise is the foundation of the lexical restructuring model (LRM; Metsala & Walley, 1998; Walley, 1993), which has gained empirical support from converging sources, including for example, studies of infant perception (Jusczyk et al., 1994), phonemic awareness and early reading skills (Walley et al., 2003), spoken word recognition (Metsala, 1997a), and corpora analyses (Logan, 1992). These data notwithstanding, it is relevant that there are other reports, which run counter to claims of a density advantage in language acquisition (e.g., Dollaghan, 1994; Gierut & Morrisette, 1998; Metsala, 1997b). Although discrepancies in results may be traced to differences in methodology, task, modes of responding, or age, there is one critical issue that has largely been ignored with respect to the LRM. Specifically, the LRM predicts that dense neighborhood structure goes hand in hand with more complex or advanced phonological representations. To our knowledge, there is just one study that has attempted to establish whether phonological complexity is realized relative to neighborhood density. In that study, Smith, McGregor, and DeMille (2006) evaluated typical children, who varied in their lexical proclivity such that some presented with age-expected vocabulary

skills and others, with precocious abilities. By the LRM, those with precocious lexical skills/denser neighborhoods would be expected to also have more advanced phonologies. Results showed, however, that despite differences in vocabulary size, children were strikingly similar in their acquisition and use of the ambient phonology (for similar findings for children with phonological delays, see Morrisette, McAllister, & Gierut, 2009). Thus, the predictions of the LRM, particularly as associated with phonological structure, were not fully borne out. This study is especially germane because it raises questions about whether dense neighborhood structure does motivate elaboration or expansion of the phonological system.

It is possible to probe this issue in complementary ways by turning to children with phonological delays enrolled in treatment and asking whether treatment of an erred sound in words from dense neighborhoods prompts greater phonological learning. This question was put to test in a series of studies (Gierut, Morrisette, & Champion, 1999; Morrisette & Gierut, 2002). The general finding was that treatment of words from dense neighborhoods resulted in little to no phonological generalization. Although children learned to accurately produce the treated sound in dense words during instruction, they did not also evidence transfer of learning to the treated or other untreated sounds that were in error. Dense words did not promote the expected lexical restructuring and expansion of the sound system. By comparison, in that same work, treatment of an erred sound in words from sparse neighborhoods did facilitate transfer across the sound system. Note that these treatment results mirror the inhibitory patterns associated with dense neighborhoods and the facilitatory patterns associated with sparse neighborhoods as reported for adults. Neighborhood density then is one word-level variable that bears on phonological learning in treatment of the population of study, with sparse (not dense) neighborhood structure having positive consequences for the phonology.

WORD FREQUENCY AND PHONOLOGICAL TREATMENT

Although neighborhood density may serve to organize the lexicon, it has also been suggested that each word in a neighborhood has additional properties, which may further impact the process of spoken word recognition. Word frequency, in particular, exerts especially robust effects on behavior. High frequency words that occur often in language facilitate the speed and accuracy of responding in studies of adults, children, and infants (e.g., Jescheniak & Levelt, 1994; Metsala, 1997a; Oldfield & Wingfield, 1965; Plunkett & Marchman, 1991). This pattern also holds for children with phonological delays enrolled in treatment. When an erred sound is taught in high frequency words, children show systemwide phonological advances, with improved production accuracy across the board to treated and untreated sounds (Gierut et al., 1999; Leonard & Ritterman, 1971; Morrisette & Gierut, 2002; cf. Moore, Burke, & Adams, 1976).

Of relevance to the present study is that word frequency is inversely correlated with another attribute of words, namely, AoA. AoA norms estimate the relative age at which a given word enters the lexicon. Conventionally, AoA is established from subjective ratings, whereby groups of participants are asked to judge when they might have acquired a given set of test words (Gilhooly & Watson, 1981, and references therein). Each test word is rated, using either a 7- or 9-point scale, with

each point on the scale corresponding to an age interval. For example, Gilhooly and Logie (1980) employed a 7-point scale to document the AoA of 1,944 words. In their study a rating of 1 translated to an AoA of 0 to 2 years, whereas a rating of 7 corresponded to an AoA of 13 years or older. Their data showed that a word like “button” was early acquired, having been assigned a subjective rating of 1.92, which corresponded to the 0- to 2-year range. A word like “audit,” by contrast, was acquired later, having received a subjective rating of 6.42, which translated to the 11- to 12-year range. It is of note that subjective AoA estimates (albeit 7- or 9-point scales) have high interscale validity (Carroll & White, 1973; Gilhooly & Gilhooly, 1980; Snodgrass & Vanderwart, 1980; Walley & Metsala, 1992). It is of further relevance that AoA has demonstrable behavioral effects on spoken word recognition in adults and children (Dewhurst, Hitch, & Barry, 1998; Ellis & Morrison, 1998; Gerhand & Barry, 1999a, 1999b; Morrison & Ellis, 1995; Walley & Metsala, 1992). Specifically, a consistent finding is that early acquired words facilitate the speed and accuracy of responding relative to later acquired words.

This observation has attracted theoretical interest in light of the known AoA–frequency correlation, with various hypotheses having been advanced about the locus of the AoA effects. Consider that early acquired words have presumably resided in the lexicon the longest; consequently, they have greater cumulative and experienced frequency (Lewis, 1999a). Early acquired words are likely to be highly familiar, requiring less extensive processing, with less interference during retrieval (Dewhurst et al., 1998). They might also be phonologically well specified, having undergone lexical restructuring as neighborhood density expands (Metsala, 1997a). Early acquired words are often shorter in length and, for this reason, some might argue they are easier to pronounce (Farwell, 1976). Further, early acquired words are more likely to reference concrete concepts that are readily accessible on semantic grounds, compared to later acquired words that tend to be associated with abstract concepts (Belke, Brysbaert, Meyer, & Ghyselinck, 2005). Thus, the benefits of early acquired words may be due to a number of factors associated with, but not limited to encoding, retrieval, representation, production, and/or semantic transparency. Given this, one question that arises is whether AoA is the higher order word-level property, with word frequency being its derivative (Morrison & Ellis, 1995; cf. Gerhand & Barry, 1998).

The same question may be raised about phonological learning by children with delays. Because treatment of high frequency words led to greater phonological gains for these children (Gierut et al., 1999; Morrisette & Gierut, 2002), it might be that these were simply early acquired words masquerading as high frequency items. To our knowledge, there are no reports of the effects of AoA in treatment of phonological delays. Moreover, the verdict is out on the role of AoA in typical phonological development. In typical development, AoA has been considered only with respect to the acquisition of the voice contrast in stops, as documented in acoustic studies. By one account, voicing is first learned in words that are early acquired, but only in those forms that are also of high frequency (Tyler & Edwards, 1993). By another account, voicing is first learned in later acquired words (Macken & Barton, 1980). Yet another view is that AoA has no bearing on the acquisition of voicing (Mack & Lieberman, 1985). Thus, the function of AoA in typical and atypical phonological development remains largely unknown.

The purpose of this study was to evaluate the effects of AoA on phonological learning. Toward this end, children with phonological delays were differentially enrolled in treatment of erred sounds that were taught in either early or late acquired words. Treated words were further dissociated in terms of word frequency to unpack the correlation with AoA. Following from the literature, one prediction is that treatment of early acquired words will facilitate phonological learning. This is consistent with the effect of treating high frequency words and the observed AoA–frequency correlation. An alternate prediction is that the effects of AoA will not align with those of word frequency. It is possible that, by disambiguating AoA and frequency in treatment as planned, the variables will exert different effects on children’s phonological learning. The results to emerge will provide preliminary data about the bearing of AoA on phonological learning, with potential applied consequences for treatment of children with phonological delays.

Participants

Ten children were recruited by public announcement to area schools, day care facilities, and early childhood programs. To participate, children had to be identified as having a phonological delay, characterized by a reduced consonantal inventory. This was established by performance 1 standard deviation below the normative mean reported for age- and gender-matched peers on the Goldman–Fristoe Test of Articulation, Second Edition (Goldman & Fristoe, 2000). In addition, children had to produce at least five target English sounds with 0% accuracy across contexts as sampled on the Phonological Knowledge Protocol (PKP; Gierut, 1985). This probe samples all target English sounds in each relevant word position in multiple exemplars, and provides for the elicitation of minimal pairs and morphophonemic alternations as evidence of the phonemic status of sounds in a child’s phonological system. Probe items were elicited using a spontaneous picture-naming task, with a child’s responses audiorecorded and phonetically transcribed using narrow notation of the IPA. Transcribed responses were used to compute percentages of accurate production, and submitted to standard phonological analyses to establish each child’s phonetic and phonemic inventories, distribution of sounds, and rule-governed alternations (Dinnsen, 1984).

In addition to a reduced consonantal inventory, inclusionary criteria for participation required that children be within 3 and 6 years of age, and perform within typical limits on a battery of diagnostic tests, which evaluated hearing acuity (American Speech–Language–Hearing Association, 1997), oral–motor structure and function (Robbins & Klee, 1987), intelligence (Roid & Miller, 1997), working memory (Kirk, McCarthy, & Kirk, 1968), expressive and receptive vocabulary (Dunn & Dunn, 1997; Williams, 1997) and language (Wiig, Secord, & Semel, 1992). The exclusionary criteria included literacy, bilingualism, and concurrent enrollment in intervention for speech and/or language disorders.

In all, two boys and eight girls qualified for participation. Their mean age was 56.6 months (range = 46–71 months). Table 1 shows the Goldman–Fristoe Test of Articulation, Second Edition, scores for each child, along with the sounds excluded from individual phonemic inventories and experimental assignments. Table 2 reports children’s performance on other inclusionary measures.

Table 1. *Children's phonemic inventories and experimental assignment*

Participant	GFTA-2 ^a	Sounds Excluded	Treatment Condition	Treated Sound
1	49	ŋ k g θ ð s z ʃ r	Early-high	k
2	74	v θ ð ʃ r	Early-high	ʃ
3	67	k g θ ð ʃ ʧ ʤ l r	Early-low	k
4	50	ŋ k g θ ð ʃ ʧ ʤ r	Early-low	ʃ
5	78	f v θ ð ʃ	Early-low	f
6	69	ŋ k g θ ð ʃ ʤ r	Late-high	k
7	57	s z ʃ ʧ l r	Late-high	ʃ
8	45	ŋ k g θ ð ʧ l r	Late-high	ʧ
9	<40	k g θ ð r	Late-low	k
10	76	θ ð ʃ ʧ r j	Late-low	ʃ

^aStandard score on the Goldman–Fristoe Test of Articulation—Second Edition (Goldman & Fristoe, 2000).

Table 2. *Children's performance on other inclusionary measures*

Participant	Hearing ^a	Oral Motor ^b		Leiter ^c	PPVT ^d	EVT ^e	CELF ^f	ITPA ^g
		Structure	Function					
1	Passed	24	99	123	99	109	101	30
2	Passed	24	104	110	95	100	107	32
3	Passed	24	108	102	103	94	100	28
4	Passed	24	102	95	93	101	96	30
5	Passed	24	102	122	131	124	119	36
6	Passed	24	103	128	113	118	105	37
7	Passed	24	106	91	93	99	94	36
8	Passed	23	102	88	103	98	103	33
9	Passed	24	101	131	106	98	98	31
10	Passed	24	104	113	108	114	103	30

^aAudiological screening passed at 20 dB HL for 1000, 2000, and 4000 Hz in accordance with ASHA standards (ASHA, 1997).

^bClinical assessment of oropharyngeal motor development in young children (Robbins & Klee, 1987). Structure score range for ages 46–71 months is 20–24; function score range for ages 46–71 months is 99–112.

^cStandard score on Leiter International Performance Scale—Revised (Roid & Miller, 1997).

^dStandard score on Peabody Picture Vocabulary Test (3rd ed., Dunn & Dunn, 1997).

^eStandard score on Expressive Vocabulary Test (Williams, 1997).

^fStandard score on Clinical Evaluation of Language Fundamentals—Preschool (Wiig et al., 1992).

^gScaled score ($M = 30$, $SD = 6$) on Illinois Test of Psycholinguistic Ability—Auditory sequential memory subtest (Kirk et al., 1968).

EXPERIMENTAL DESIGN

A staggered multiple baseline (MBL) across subjects design was used in the experimental manipulation of treatment (McReynolds & Kearns, 1983). The MBL involves a period of no treatment followed by treatment. The no treatment or baseline phase is incremented as successive children are enrolled, such that the number of baselines increases by one with each leg of the design. The underlying premise is that a given child's performance remains stable during baseline and changes only with the instatement of treatment, establishing a cause-effect relationship. The ongoing stability of baselines across successive legs of the design is taken as demonstration that changes in performance are not attributable to maturation. If maturation were at work, then the time-lagged baselines of successive children in the MBL would likewise evidence improvement because these children were afforded time to mature. Successive legs of the MBL also provide for direct and systematic replications across children as a reflection of the generalizability of treatment effects.

As applied in this study, the independent variable was treatment of one erred sound in stimulus words that systematically varied in AoA value; these are described in detail below. AoA was dissociated from word frequency, such that there were four experimental conditions: treatment of early versus late acquired words that were either of high versus low frequency (early-high, early-low, late-high, late-low). Two children each were assigned to the correlated conditions, early-high and late-low, where AoA and frequency dovetailed. Three other children each were assigned to the noncorrelated conditions, early-low and late-high. The latter were of particular interest because they held the potential to disambiguate the correlation between AoA and word frequency; hence, the increased number of participants (replications). The dependent variable was generalization, defined as the percentage accuracy of treated and untreated sounds excluded from a child's pretreatment phonemic inventory. Generalization was measured using the PKP, with those sounds evidencing positive gains pre- to posttreatment being entered into data analysis.

Direct and systematic replications were incorporated into the design to establish generalizability (McReynolds & Kearns, 1983). This is illustrated in Table 3, which provides an overview of the experimental design. Direct replications were planned in the orthogonal variation of AoA and word frequency, enrolling two or three children. Systematic replications were planned by blocking experimental variables, in evaluation of AoA independent of word frequency, and of word frequency independent of AoA. Direct and systematic replications were also accommodated in the sound that was treated based on projected orders of acquisition, as outlined below.

Treated stimuli

Each child was taught one erred sound in the initial position of six stimulus words that were specifically selected to align with the assigned experimental condition. The treated sound was pseudorandomly selected based on observed gaps in a given child's presenting phonemic inventory; however, the treated sound was limited to either /k/, /j/, /f/, /tʃ/. These are among the mid- and late-8 phonemes,

Table 3. *Overview of the experimental design*

Word Frequency	AoA	
	Early Acquired	Late Acquired
High	Mid-8	Mid-8
	Late-8	Late-8
Low	Mid-8	Mid-8
	Late-8	Late-8

as defined by a classification scheme (Shriberg, Kwiatkowski, & Gruber, 1994) that organizes sounds into sets of eight based on their relative order of emergence in children's inventories. Treated sounds were chosen from the mid- and late-8 categories because these are commonly lacking in the inventories of children with phonological delays (Ingram, 1997; Shriberg et al., 1994). By varying the treated sound across children and conditions in this way, direct and systematic replications based on mid- versus late-8 phonemes were afforded. This also controlled for spurious effects of the treated sound on phonological learning. By incorporating more than one treated sound into the overall design, it was possible to determine that learning was due to AoA and not the treated sound per se.

Stimulus words used in treatment were chosen based on their AoA and word frequency values. AoA was defined in reference to Gilhooly and Logie (1980) and Bird, Franklin, and Howard (2001). It was necessary to consult two AoA corpora to yield an adequate set of items that were picturable, child-appropriate, and crucially distinct in their word frequency values. As was noted, the Gilhooly and Logie corpus comprises 1,944 words, with Bird et al. (2001) consisting of 2,694 words. The corpora are both based on 7-point rating scales, and shown to be comparable (Bird et al., 2001). In this study, the mean AoA of early acquired words was 2 (estimated age of 3–4 years). The mean AoA of late acquired words was 4 (estimated age of 7–8 years). Early and late AoA categories were confirmed as statistically independent by *t* test analysis, $t(58) = -9.16, p = .000$. Word frequency was operationalized relative to Kučera and Francis (1967). The mean raw frequency was 291 occurrences per million for high frequency items and 22 occurrences per million for low frequency items. These were also statistically independent categories, $t(58) = 2.84, p = .006$.

Treated stimuli were weighted to partially favor words from dense neighborhoods, where density was defined as 10 or more neighbors (Luce, 1986). The partial weighting of dense neighborhood structure was intended to balance the observed inhibitory effects of density on phonological generalization (Morrisette & Gierut, 2002) against the apparent facilitatory effects of density that are associated with lexical restructuring (Walley, 1993). Across conditions, the mean density of treated words was 13. Within a condition, the ratio of dense to sparse words was 1.4:1. Thus, all children were presented with a mix of words from sparse and dense neighborhoods in an effort to potentially promote phonological generalization and/or lexical restructuring.

Table 4. *Characteristics of treated stimuli*

	Treatment Conditions				<i>F</i> ^a	<i>p</i>
	Early-High	Early-Low	Late-High	Late-Low		
Imageability	441.43	484.33	394.44	588.75	2.927	.06
Familiarity	6.96	6.98	6.94	6.93	0.463	.71
Neighborhood frequency	160.72	189.97	107.91	41.88	1.649	.19
Sum of segment frequency ^b	0.59	0.54	0.15	0.27	0.572	.64
Sum of biphone frequency ^b	0.47	0.20	0.35	-0.02	0.361	.78
Word length	3.33	3.44	3.83	3.83	1.198	.32

^aDegrees of freedom for imageability (3, 19); all others (3, 56).

^bSum of segment frequency and sum of biphone frequency values represent *z*-score transformations to control for word length (Storkel, 2004).

In addition, treated stimuli were further equated on a range of complementary variables to minimize potential extraneous influences on phonological learning. As in Table 4, the following were controlled across experimental conditions: imageability, familiarity, neighborhood frequency, phonotactic probability (i.e., sums of segment frequency and biphone frequency), and word length. Of these, imageability warrants comment due to debates about whether AoA has a semantic or phonological origin (Brysbaert, Van Wijnendaele, & De Deyne, 2000). The imageability of words is thought to reflect their semantics (Snodgrass & Vanderwart, 1980); therefore, by controlling imageability, the semantics of the treated stimuli would be held constant so as to focus on the phonological side of AoA. One-way analyses of variance showed no statistically significant differences in the variables that were controlled, as reported in Table 4. Further, a chi-square test showed no statistically reliable difference in the distribution of treated stimuli by syntactic category across early versus late acquired conditions, $\chi^2(1, N = 60) = 4.31, p > .05$.

Treatment procedures

Children were provided treatment three times weekly in 1-hr individual sessions. Treatment followed a conventional protocol (Gierut, 2008) consisting of two phases. The first phase involved imitation of the one treated sound in the treated words. During imitation, a child viewed a picture of the treated word on a computer screen; the clinician said the depicted word; then the child was to repeat this word. Feedback about the accuracy of responding was provided on a continuous schedule. In the case of inaccurate productions, the clinician provided instructions about articulatory placement, and an additional model of the form was presented. Imitation continued until a child achieved 75% accuracy of production of the treated sound in the treated words over two consecutive sessions or until seven

total sessions were completed, whichever occurred first. Following this, treatment shifted to spontaneous naming. The child again viewed the picture of the treated word on a computer screen, but was to independently name the item with accurate production of the treated sound. Feedback and instruction were provided in the same manner outlined for the imitation phase. The spontaneous phase continued until a child achieved 90% accuracy of production over 3 consecutive sessions or until 12 total sessions were completed, whichever occurred first. The treated sound and stimuli remained constant throughout treatment.

During treatment, generalization learning was measured using the full PKP, along with various subsets of that probe. The full PKP was administered at baseline, phase shift from imitation to spontaneous production, and again immediately posttreatment. During treatment, only those sounds excluded from a given child's inventory were sampled, with subset probes being administered on a variable ratio schedule averaging two sessions.

Data analysis and transcription reliability

Day to day accuracy in treatment was evaluated to determine advancement through the treatment program, and to demonstrate that learning had taken place. Across conditions, all children evidenced learning of the treated sound. Mean accuracy of production of the treated sound in the treated stimuli was 89% during the imitation phase, and 98% during the spontaneous phase of instruction. Performance for children assigned to the early AoA conditions ranged from 85% to 100% accuracy and for those assigned to the late AoA conditions ranged from 73% to 100% accuracy. This served to document that phonological learning occurred under the different AoA conditions, from which children were left free to generalize in keeping with the dependent variable.

Generalization was computed as percentage accuracies based on the PKP performance of each child. Recall that generalization was the primary dependent measure, and that all sounds excluded from a given child's pretreatment inventory were monitored in all relevant word positions for production accuracy. This included the treated sound, as well as other untreated sounds in error at baseline. These data were derived longitudinally from phonetic transcriptions of the PKP. To ensure reliability of transcription, an estimate of interjudge agreement was obtained. Two trained listeners independently transcribed approximately 10% of the PKP data from each child using narrow notation of the IPA. Consonant transcriptions were compared point to point, with mean interjudge agreement being 93% based on 2,982 consonants transcribed.

Generalization data from each child were then examined from qualitative and quantitative perspectives, following procedures established in the literature (Morrisette & Gierut, 2002, and references therein), with one new addition. Briefly, the qualitative description of generalization applied the accepted criterion level of 10% or greater mean generalization accuracy (Elbert, Dinnsen, & Powell, 1984) as the operational definition of learning. For each experimental condition, generalization learning was examined to see if the 10% criterion was met. The resulting description afforded initial insight to which treatment conditions might have induced generalization learning.

This was followed by two complementary quantitative analyses to determine whether the treatment conditions also resulted in statistically reliable generalization and to establish the relative magnitude of the generalization gains. The rank procedure for replicated AB MBL design was applied consistent with Morrisette and Gierut (2002) to establish generalization by condition as a statistically reliable effect. In short, this procedure makes use of the combined Sign Test (Marascuilo & Serlin, 1988) by combining behaviors across subjects for increased power. Mean generalization accuracy of the treated sound, untreated (erred) sounds from the same manner class as the treated sound, and untreated (erred) sounds from different manner classes than the treated sound were factored into analysis for each child. For each factor, mean baseline performance and longitudinal generalization were ranked, with the difference between ranks totaled by experimental condition; this takes into account the range of variation observed within a condition. A binomial distribution ($p = .05$) was then used to determine whether the resulting generalization was statistically reliable by condition. It should be noted that the number of factors that went into analysis varied due to the number of children enrolled (i.e., two vs. three replicants). The opportunity for different types of generalization also varied given children's presenting inventories (e.g., Child 5 only excluded fricatives as shown in Table 1; hence, generalization across manner classes could not be instantiated).

In addition to the rank procedure, quantitative analyses included the calculation of effect size for single-subject experimental design. To our knowledge, effect size has not been reported in prior single-subject studies of phonological treatment and is thus a novel extension. For this reason, the use of effect size for single-subject experimental design warrants further explanation.

Effect size computations are being increasingly advocated for single-subject designs to quantify treatment efficacy, with an eye toward contributing to meta-analyses of small n studies (Schlosser, 2005). Effect size captures the magnitude of learning that takes place under a given treatment condition. Effect size values are then compared across conditions to identify which is relatively most efficacious. Both regression and nonregression formulas have been evaluated to determine their applicability to single-subject design. A consensus is that nonregression techniques are more conservative and tend not to overestimate the effects of treatment (Campbell, 2004).

Within the recommended family of nonregression techniques, different effect size metrics are likewise available. In evaluation of these, the standard mean difference (SMD) has emerged as a preferred analysis (Olive & Smith, 2005). To compute SMD, the mean of the baseline data is established, along with the mean of the (probe) data collected during treatment. Then, the difference between baseline and probe data is obtained, and divided by the standard deviation of the baseline to achieve the effect size d . This takes into account baseline variability within a condition (Busk & Serlin, 1992). There are two variants of SMD, one that utilizes the first and final three points of baseline and probe data, respectively (SMD₃), and one that utilizes all baseline and probe data (SMD_{all}). Although both are acceptable, the latter is preferred because it employs the richest set of data, is easy to compute, is most readily interpretable, and results in a conservative estimate of gain.

Table 5. Results of qualitative and quantitative analyses by condition

	Mean Generalization Accuracy	Rank Procedure		Effect Size
		<i>n/N</i> ^a	<i>p</i>	<i>d</i>
Early-high	4%	1/5	.16	2.81
Early-low	5%	4/5	.16	3.66
Late-high	12%	7/7	.01	11.41
Late-low	13%	5/5	.03	16.58

^a*n*, Number of factors evidencing positive change; *N*, number of possible factors probed for positive change (i.e., number of subjects × treated sound and other untreated sounds from the same and different manner classes; see Morrisette & Gierut, 2002, p. 150).

The interpretation of *d* then follows from empirically generated standards about the magnitude of effect for particular treatments. Although benchmarks for the *d* statistic have been established for between group designs, it is inappropriate to extend these to other contexts (Cohen, 1988). Consequently, single-subject applications must begin to generate standards to assess whether specific treatments yield small, medium, or large effect sizes. Thus far, tentative benchmarks have been identified for syntactic production and lexical retrieval treatments applied to aphasia (Beeson & Robey, 2006). For syntactic treatment, effect sizes of 6.0, 12.0, and 18.0 correspond to small, medium, and large effects, respectively. For lexical retrieval treatments, 4.0, 7.0, and 10.1 have been advanced as small, medium, and large effect sizes, respectively. To our knowledge, there are no studies of phonological treatment per se that have offered guidelines for interpretation of magnitude of effect; this study thus will be a first step in this direction.

For purposes of the present study, SMD_{all} was used in data analysis to establish the relative magnitude of generalization that obtained. SMD_{all} was computed for the orthogonal manipulations of AoA relative to word frequency to document *d* as the effect size by condition. In addition, independent effect sizes were computed for AoA (early vs. late), word frequency (high vs. low), and age of sound acquisition (mid- vs. late-8), following from the planned direct and systematic replications that were incorporated into the experimental design (Table 3).

RESULTS

The primary research question assessed the effects of AoA on children’s phonological generalization, when orthogonally varied with word frequency. Table 5 summarizes the results of the qualitative and quantitative analyses for each of the four experimental conditions.

Beginning with the descriptive analysis, Table 5 shows that treatment of a sound in early acquired words (albeit high or low frequency) resulted in less than 10% mean generalization accuracy, as the operational definition of phonological learning. Children assigned to the early-high AoA-frequency condition achieved

an average of 4% accuracy relative to baseline. Those in the early-low AoA-frequency condition performed much the same, achieving 5% accuracy. These patterns of learning contrasted with treatment of late acquired words. In these conditions, children met the 10% criterion, averaging 12% and 13% accuracy for the late-high and late-low AoA-frequency conditions, respectively. On qualitative grounds, it appeared that treatment of late acquired words induced phonological generalization, whereas early acquired words did not.

Quantitative analyses using the rank procedure for replicated AB MBL designs confirmed this observation. Table 5 shows that neither of the early AoA treatment conditions resulted in statistically significant generalization effects ($p > .05$). By comparison, both late AoA treatment conditions yielded statistically reliable generalization effects ($p < .03$).

The pattern of results was further corroborated by the relative magnitude of generalization associated with each condition. The d values reported in Table 5 show that relatively greater effect sizes obtained for the two late AoA conditions. Late acquired words led to greater phonological gains, with d values of 11.4 and 16.6, respectively, for high and low frequency treated stimuli. This contrasted with early acquired words, where d equaled 2.8 and 3.7, respectively, for high and low frequency treated stimuli. It is of note that the AoA effects appeared to be independent of word frequency. This can be seen in the comparable d values for the two early AoA conditions, high and low frequency, and likewise, the two late AoA conditions, high and low frequency. This aligns with developmental reports that AoA and word frequency are distinct lexical variables, which can be disambiguated (Garlock, Walley, & Metsala, 2001), and further hints that the two may exert separate influences on phonological learning. The planned direct and systematic replications afford an opportunity to explore this point further in evaluation of the independent effects of each variable.

In this regard, the respective contributions of each independent variable (AoA, word frequency, and age of sound acquisition) were again examined using effect size data. Here, effect sizes were blocked and averaged for a given variable, while ignoring its other dimensions (following Olive & Smith, 2005; see also Dewhurst et al., 1998). To illustrate, with reference to Table 3, the unique contribution of AoA was determined by averaging d values for all children assigned to early AoA conditions, and this was compared to the average d for those assigned to late AoA conditions. The treated words' frequency and mid-/late-8 status of the treated sounds were set aside in the blocking. Similar calculations were developed for the effects of word frequency. Again, with reference to Table 3, high frequency conditions were collapsed and compared to low frequency conditions, while ignoring AoA and age of sound acquisition. For completeness, the effect of age of sound acquisition was also examined, with mid-8 conditions merged and compared to late-8 conditions. Each of the independent comparisons is plotted in Figure 1, Figure 2, and Figure 3, respectively.

Figure 1 shows that the independent contribution of AoA to phonological generalization is consistent with the aforementioned orthogonal comparisons: a greater effect size was associated with late acquired words ($d = 13.5$) compared to early acquired words ($d = 3.3$). The relevant and added point that is established here relates to the relative magnitude of gain that accrues solely from AoA, aside from

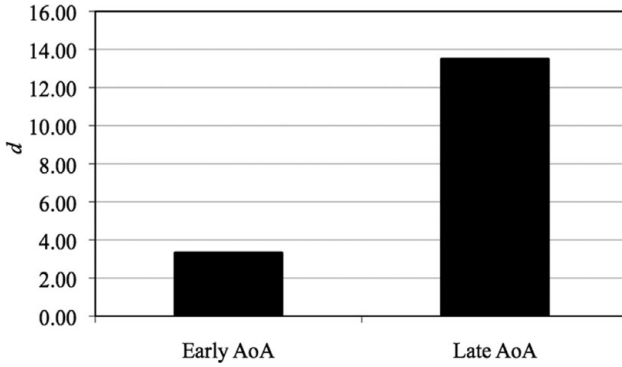


Figure 1. Effect sizes for the comparison of early versus late acquired words. AoA, age of acquisition.

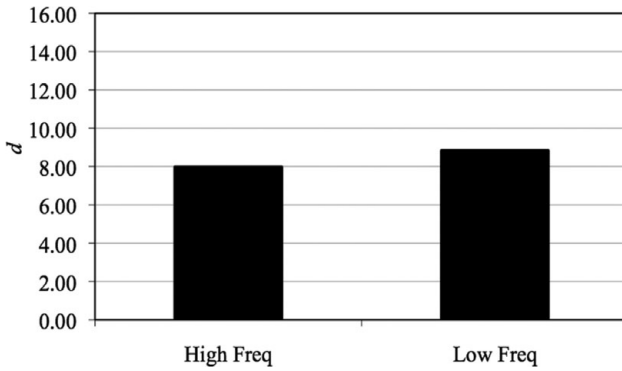


Figure 2. Effect sizes for the comparison of high versus low frequency words.

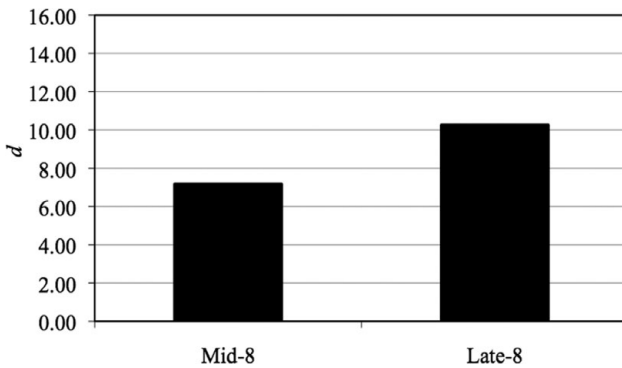


Figure 3. Effect sizes for the comparison of mid- versus late-8 sounds.

contributions of word frequency or age of sound acquisition. Of significance, the effect size of late acquired words was four times that of early acquired words, supporting the former as having the most impact on phonological learning.

The contribution of word frequency to phonological generalization, when AoA and age of sound acquisition were removed from consideration, is shown in Figure 2. It can be seen that the magnitude of gain was essentially equivalent for high and low frequency words ($d = 8.0$ and 8.8 , respectively). There was no apparent differential generalization associated with word frequency when AoA and age of sound acquisition were subbed out.

The contribution of age of sound acquisition, separate from AoA and word frequency, is shown in Figure 3. The effect size associated with treatment of late-8 sounds ($d = 10.25$) exceeded that of mid-8 sounds ($d = 7.17$), such that the magnitude of gain was approximately 1.4 times greater for late-8 sounds relative to mid-8 sounds.

DISCUSSION

Taken together, the orthogonal and independent analyses identified AoA as a property of words that contributes to phonological generalization in treatment of children with phonological delays. More specifically, late acquired words promoted greater generalization than early acquired words in magnitude of 4:1. This finding is striking in light of literature that has shown an advantage for early acquired words in spoken word recognition and processing by children and adults (Garlock et al., 2001). Thus, one question for discussion is why there was a late AoA advantage for children with phonological delays in treatment.

The results also revealed that AoA was distinguishable from word frequency. Although AoA resulted in differential generalization, word frequency did not. Exposure to high or low frequency words in treatment led to essentially the same magnitude of gain. This finding also appears to stand apart from the literature, which reports that high frequency words in treatment promote greater generalization for children with phonological delays (Morrisette & Gierut, 2002). Moreover, high frequency words facilitate spoken word recognition and processing by infants, children, and adults (Metsala, 1997a; Oldfield & Wingfield, 1965; Plunkett & Marchman, 1991). Thus, a second question for discussion is why a high frequency advantage was not also apparent in the present study.

A final finding was that children benefited from treatment of late-8 phonemes relative to mid-8 phonemes in magnitude of 1.4:1. This is in keeping with, and serves to replicate the existing literature (Gierut, Morrisette, Hughes, & Rowland, 1996). For completeness, the discussion will also consider the interface of age of sound acquisition with AoA as a possible direction for future research.

Late AoA benefits phonological learning

For insight to the discrepancy between our finding of a late AoA advantage for phonological learning versus other reports of an early AoA advantage for other behavioral responses, we return to general accounts of AoA effects, as outlined in the introduction. Although a range of factors has been suggested as causal to

the early AoA advantage, perhaps the most widely applied proposal is the phonological completeness hypothesis (PCH; Brown & Watson, 1987). The PCH posits that early acquired words are initially stored as, and remain robust phonological wholes throughout the life span (Brown & Watson, 1987, p. 214; Ferguson & Farwell, 1975; cf. “structural–residual effects,” Metsala, 1997a). Holistic storage presumably enables the facilitating effect of early acquired words because they are likewise, retrieved holistically. However, as vocabulary size increases, a child must find more economical ways to store lexical items. One way to economize is to segment words into their subconstituent phonological structures (e.g., features, phonemes, syllables). This allows words that share the same phonological properties to be represented by the same linguistic units, with three consequences. First, late acquired words are assumed to have greater phonological structure than early acquired words due to the segmentation strategy for lexical storage. Second, a child is presumed to have knowledge of the subconstituent structure of late acquired words, whereas the same is not true for early acquired (holistic) forms. Third, although the segmentation tack may improve lexical storage, it has costs on processing because each time a word is retrieved, its subconstituent parts must be newly assembled. Assembly takes time, resulting in the slowed effects that have been reported for late acquired words. It is of mention that, under the PCH, the distinction between early and late acquired words is attributed to a developmental shift in the strategy that a learner uses to differentially encode and retrieve the respective items (for a similar proposal, see also Aitchison & Straf, 1981). It is the strategy that is modified, not the internal phonological properties of the lexical representations themselves. This contrasts with the LRM, which purports that lexical representations undergo change, becoming increasingly refined in phonological substance (for further discussion, see Monaghan & Ellis, 2002; Walley et al., 2003).

If the PCH is correct, then one possible prediction is that treatment of an erred sound in late acquired words may have been responsible for the apparent shift in strategy, from holistic to subconstituent storage of lexical items. Consider that when late AoA words were taught to the children of this study their subconstituent structure may have been brought to the forefront. On each and every trial in treatment, a child would have had to utilize syllables, segments, and/or features to access, recognize, extract, assemble, and produce the late acquired treated words. This level of analysis may have promoted generalization to like phonological units, resulting in the late AoA advantage. The subconstituent structure that is apparently associated with late acquired forms, coupled with the extraction of that structure for purposes of word assembly, may have aided phonological generalization (for similar a line of reasoning, see Monaghan & Ellis, 2002; Morris, 1981). If this premise is correct, then the present data from phonological learning may be added to those from spoken word recognition in support of the PCH, while allowing for distinct functions of early versus late acquired words depending upon the task at hand.

Although this explanation may be plausible, future studies will need to further verify the PCH as a viable account of the late AoA advantage in phonological learning. To do this, it may be necessary to collect reaction time and/or kinematic data as complements to generalization data. Recall that the PCH posits subconstituent structure for late acquired words, but this structure must be assembled on each and every encounter with those forms. It is possible that, although late

acquired words promote phonological accuracy, production of those forms may be slowed. Combined speed and accuracy data from production would be one way to document the presumed slowing effects of late acquired forms. Extensions of this sort would allow for the full range of assumptions inherent to the PCH to be tested within the domain of phonology.

With respect to typical phonological development, the finding of a late AoA advantage for children with phonological delays is consistent with Macken and Barton's (1980) proposal that the last words learned are the first to undergo phonological change. Although the present study did not trace the AoA of the words that generalized, it did reveal, like Macken and Barton (1980), that late acquired words advance the course of phonological acquisition. This notwithstanding, questions remain about the interpretation and application of AoA relative to a given child's lexicon. Consistent with the broader literature, AoA was defined in this study based on external judgments and corpora. Although AoA has been shown to be a statistically reliable parameter, it is not possible to know exactly what the AoA status of a given word was for a given child. This could only have been determined by tracking a child's longitudinal acquisition of words beginning with the first outputs of his or her expressive lexicon (Jorm, 1991). Even under ideal circumstances, the precise delineation of word acquisition would be a formidable task, subject to omission, error, and best guess estimates.

It is important to recognize that, whenever a statistical property of language has been applied to acquisition, similar concerns have been raised (e.g., Dollaghan, 1994). At issue is whether available corpora provide an accurate reflection of a developing lexicon. Arguments in support of the appropriateness of such extensions have tended to appeal to the normalization of standard distributions (Kelly & Martin, 1994), correlations to show the equivalence of corpora (Burgess & Livesay, 1998), and/or behavioral demonstrations of comparability across corpora (Gierut & Dale, 2007). Nonetheless, in future studies of AoA, it may be worthwhile to obtain a preliminary assessment of whether a child has a given word in his or her lexicon. This might be revealed in a forced-choice receptive identification task to globally approximate a distinction between early and late acquired words. Words identified by a child might be deemed "early" acquired, whereas those that are not might be thought of as "late" acquired. The identification data would then have to be compared and confirmed against AoA norms. Dovetailing the data would be necessary to eliminate the possibility that word identification corresponds only with word familiarity, and not AoA. Corroborating data of this sort may help ground AoA within the context of a given child's lexicon, and this in turn may help refine the interpretation of results.

The present findings have possible implications for phonological treatment and evidence-based practice. A recommendation that emerges from this research is that efficacy may be enhanced by treatment of erred sounds in late acquired words. This recommendation applies only if the goal of treatment is to promote systemwide phonological generalization to treated and untreated sounds and contexts, as was the case here. Although systemwide gains are the ultimate outcome of treatment (Olswang, 1998), there are at least three related questions that may be entertained. One relates to the effect of late acquired words in session by session learning as treatment is being delivered, as a complement to this focus on generalization

learning. A second involves developing a more detailed description of generalization to establish whether late acquired words promote differential learning in particular facets of the phonology, as in a comparison of within- and across-class generalization. A third bears on the efficiency of treating late acquired words, as measured by the time in treatment to induce generalization. With these pieces in place, it may be possible to delineate the full range of benefits that accrue from exposure to late acquired words in phonological treatment. There are other pluses associated with work along these lines, namely, sample size would be increased and a broader range of error patterns would be considered. With respect to evidence-based practice, the present findings contribute a first application of effect size computations to phonological treatment. Although preliminary, the data reported offer early estimates of the potential size of the learning effects that may accrue with phonological treatment. The documentation of effect size in future research may ultimately yield guidelines that delineate what is a small, medium, or large effect relative to phonological intervention.

AoA obscures word frequency effects

In this study, the absence of differential word frequency effects on children's phonological generalization may seem puzzling at first glance, given the robustness of high frequency words for processing, performance, and learning. There is one body of literature, however, that has largely failed to show the benefits of word frequency on performance. That work is precisely associated with experimental evaluations of AoA (e.g., Belke et al., 2005; Carroll & White, 1973; Ellis & Morrison, 1998; Garlock et al., 2001; Morrison & Ellis, 1995; Morrison, Ellis, & Quinlan, 1992). In particular, when the correlation between word frequency and AoA is controlled, AoA has emerged as the predictor variable. This effect has been observed for a variety of tasks (e.g., object naming, picture recognition, mispronunciation detection, vocabulary monitoring), with the results of the present study on children's phonological generalization being added to that pool of data.

There have been various debates about, and hypotheses in account of the absent word frequency effects in such studies. For instance, Morrison and colleagues (1992; Morrison, Chappell, & Ellis, 1997) argued that prior frequency effects have been given undue importance because of the failure to control for AoA. When these investigators reanalyzed the classic data of Oldfield and Wingfield (1965), they found that the effects of word frequency were completely eliminated, simply by adding AoA as an independent variable. Others have argued that the hold on word frequency is due to its necessity in computational approaches to modeling lexical processing, which depend on frequency in their algorithms (Ellis & Morrison, 1998). Aside from the debates (Lewis, 1999a, 1999b; Moore, Valentine, & Turner, 1999), there are at least two hypotheses that have been put forth to explain why word frequency effects may be diminished in the context of AoA. Both appeal to the potentially different roles that AoA and word frequency may play. One hypothesis is that AoA most closely aligns with the lexical representation itself, affirming its structure and encoding, whereas word frequency impacts the strength of the connections between representations (Ellis & Morrison, 1998).

Another hypothesis, stemming from developmental work, suggests similarly that AoA exposes the phonological quality and substance of lexical representations (Garlock et al., 2001). In contrast, word frequency is presumed to affect spoken word recognition in the latest stages of processing, long after lexical access. Note that both proposals assign a phonological role to AoA, but not frequency. This observation is interesting in light of recent research by Tainturier, Johnson, Tamminen, and Thierry (2003). These investigators made use of event-related potential and behavioral data to explore whether the AoA effect stems from the phonological makeup of words or their semantic meaning. It was hypothesized that if AoA had a phonological origin, then the P300 component would be affected, but if the origin were semantic, then the N400 component would be affected in lexical decision and semantic categorization tasks, respectively. Of importance, word frequency impacts both the P300 and N400 components, making it possible to dissociate the AoA–frequency correlation. Results showed only P300 effects in lexical decision, with no N400 effects observed for semantic categorization. These data begin to lay a necessary empirical foundation in support of hypotheses that phonology is the relevant domain responsible for AoA effects. When the results of the present study are placed in the context of this broader literature, it is likely that the equivalent word frequency effects that were observed herein were due to controlling the correlation between AoA and word frequency. Like other studies, the disambiguation of AoA and word frequency in treatment may have enabled the property of AoA to emerge as relevant to phonological generalization. Further, AoA may have functioned more centrally in children’s learning, given its apparent phonological locus.

There is one additional point worth mentioning that may bear specifically on AoA relative to word frequency as applied in phonological treatment. Specifically, research has established that AoA effects may be magnified relative to word frequency during picture naming (Gilhooly & Watson, 1981, and references therein). This is potentially relevant because one way that phonological generalization is typically assessed is through the administration of structured probes. The PKP used here is a structured probe that samples children’s productions through spontaneous naming of pictures. Although prior studies have also relied on the PKP to measure generalization learning (e.g., Morrisette & Gierut, 2002), that work did not evaluate or control for AoA. The point here is that picture naming may have even further amplified the observed effects of AoA relative to word frequency in this study. Although appropriate to our experimental purpose, it may be necessary to employ complementary probe procedures in future research that do not entail picture naming. This will allow for a more detailed assessment of the contributions of different word-level variables under different sampling conditions.

Likewise, future studies may be designed to elaborate on various combinations of word-level variables. In this study, the scope was limited to the correlation of AoA and word frequency, but it may be relevant to systematically vary AoA with respect to neighborhood density (e.g., Garlock et al., 2001). Similarly, the effects of semantics (e.g., Belke et al., 2005) may be examined by orthogonally varying imageability and AoA in treatment. In this way, it may be possible to define interactions between various properties of words, even among variables that are not correlated. Theoretically, continued work along these lines may help

to verify the validity of competing proposals about the origins and/or function of AoA, particularly as evidenced in development.

AoA and age of sound acquisition

Our finding of greater generalization associated with late-8 sounds, and also late acquired words, is consistent with the literature on complexity approaches to treatment (Thompson, 2007). Following from learnability theory (Wexler & Culicover, 1980), the basic premise of this approach is that learning is facilitated when a learner is exposed to more complex structures of language, with cascading effects across the linguistic system as a whole. The present results support complexity at both sublexical (i.e., late-8 sounds) and lexical (i.e., late acquired words) levels of structure for systemwide generalization. It remains to be determined whether there is a relationship between age of sound acquisition and AoA that may further inform children's generalization learning. Future studies may explore possible precedence or additive relationships between type of treated sounds and type of treated words. Studies of this sort may contribute theoretically to a better understanding of the integration and contribution of sublexical and lexical structure to children's phonological learning. On the practical side, these extensions might help delineate stimulus priorities in the design of clinical treatment programs.

CONCLUSION

Together, the results of this study identified AoA as a property of words that contributes to phonological generalization in clinical treatment of children with phonological delays. In contrast, word frequency did not promote differential patterns of phonological learning, perhaps due to its inherent function, experimental design, and/or task used to measure the dependent variable. Learning was further affected by age of sound acquisition, with late-8 sounds inducing greater phonological generalization. These results hold clinical implications for the selection of stimuli to be used in treatment of phonological delays, namely, late acquired words and late-8 sounds. The results also suggest potentially fruitful lines of continued study that are likely to expand our understanding of how the phonology and lexicon interact for purposes of phonological acquisition. This would complement parallel strands of study that are taking place with respect to lexical acquisition (e.g., Jarvis, Merriman, Barnett, Hanba, & Van Haitzma, 2004; Magnusson, Tanenhaus, Aslin, & Dahan, 2003; Storkel, 2001). Finally, there are clear benefits that may obtain for the design and interpretation of clinical treatment in that the core characteristics of words that optimize phonological learning may be identified.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the National Institutes of Health (DC001694) to Indiana University. We thank Dan Dinnsen and members of the Learnability Project for their input and contributions to this research.

REFERENCES

- Aitchison, J., & Straf, M. (1981). Lexical storage and retrieval: A developing skill? *Linguistics*, *19*, 751–795.
- American Speech–Language–Hearing Association Audiologic Assessment Panel 1996. (1997). *Guidelines for audiologic screening*. Rockville, MD: ASHA.
- Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*, *16*, 161–169.
- Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical–semantic competition hypothesis. *Cognition*, *96*, B45–B54.
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verb and function words. *Behavior Research Methods, Instruments, and Computers*, *33*, 73–79.
- Brown, G. D., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, *15*, 208–216.
- Brysbaert, M., Van Wijnendaele, I., & DeDeyne, S. (2000). Age-of-acquisition of words is a significant variable in semantic tasks. *Acta Psychologica*, *104*, 215–226.
- Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, and Computers*, *30*, 272–277.
- 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.
- Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, *28*, 234–246.
- Carroll, J. B., & White, M. N. (1973). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*, *12*, 563–576.
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, *17*, 205–215.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Dewhurst, S. A., Hitch, G. J., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 284–298.
- Dinnsen, D. A. (1984). Methods and empirical issues in analyzing functional misarticulation. In M. Elbert, D. A. Dinnsen, & G. Weismer (Eds.), *Phonological theory and the misarticulating child* (ASHA Monographs No. 22, pp. 5–17). Rockville, MD: American Speech–Language–Hearing Association.
- Dollaghan, C. A. (1994). Children's phonological neighbourhoods: Half empty or half full? *Journal of Child Language*, *21*, 257–272.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.
- Elbert, M., Dinnsen, D. A., & Powell, T. W. (1984). On the prediction of phonologic generalization learning patterns. *Journal of Speech and Hearing Disorders*, *49*, 309–317.
- Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 515–523.
- Farwell, C. B. (1976). Some strategies in the early production of fricatives. *Papers and Reports on Child Language Development*, *12*, 97–104.
- Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*, 419–439.
- Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language*, *45*, 468–492.
- Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 267–283.
- Gerhand, S., & Barry, C. (1999a). Age-of-acquisition and frequency effects in speeded word naming. *Cognition*, *73*, B27–B36.

- Gerhand, S., & Barry, C. (1999b). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition*, 27, 592–602.
- Gierut, J. A. (1985). *On the relationship between phonological knowledge and generalization learning in misarticulating children*. Bloomington, IN: IULC Publications.
- Gierut, J. A. (2008). Fundamentals of experimental design and treatment. In D. A. Dinnsen & J. A. Gierut (Eds.), *Optimality theory, phonological acquisition and disorders* (pp. 93–118). London: Equinox.
- Gierut, J. A., & Dale, R. A. (2007). Comparability of lexical corpora: Word frequency in phonological generalization. *Clinical Linguistics & Phonetics*, 21, 423–433.
- Gierut, J. A., & Morrisette, M. L. (1998). Lexical properties in implementation of sound change. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development* (pp. 257–268). Somerville, MA: Cascadia Press.
- Gierut, J. A., Morrisette, M. L., & Champion, A. H. (1999). Lexical constraints in phonological acquisition. *Journal of Child Language*, 26, 261–294.
- Gierut, J. A., Morrisette, M. L., Hughes, M. T., & Rowland, S. (1996). Phonological treatment efficacy and developmental norms. *Language, Speech, and Hearing Services in Schools*, 27, 215–230.
- Gilhooly, K. J., & Gilhooly, M. L. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, 71, 105–110.
- Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods, Instruments and Computers*, 12, 395–427.
- Gilhooly, K. J., & Watson, F. L. (1981). Word age-of-acquisition effects: A review. *Current Psychological Reviews*, 1, 269–286.
- Goldman, R., & Fristoe, M. (2000). *Goldman–Fristoe Test of Articulation* (2nd ed.). Circle Pines, MN: American Guidance Service.
- Ingram, D. (1997). The categorization of phonological impairment. In B. Hodson & M. L. Edwards (Eds.), *Perspectives in applied phonology* (pp. 19–41). Gaithersburg, MD: Aspen.
- Jarvis, L. H., Merriman, W. E., Barnett, M., Hanba, J., & Van Haitsma, K. S. (2004). Input that contradicts young children's strategy for mapping novel words affects their phonological and semantic interpretation of other novel words. *Journal of Speech, Language, and Hearing Research*, 47, 392–406.
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 824–843.
- Jorm, A. F. (1991). The validity of word age-of-acquisition ratings: A longitudinal study of a child's word knowledge. *British Journal of Developmental Psychology*, 9, 437–444.
- Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630–645.
- Kelly, M. H., & Martin, S. (1994). Domain-general abilities applied to domain-specific tasks: Sensitivities to probabilities in perception, cognition, and language. In L. Gleitman & B. Landau (Eds.), *The acquisition of the lexicon* (pp. 105–140). Cambridge, MA: MIT Press.
- Kirk, S. A., McCarthy, J. J., & Kirk, W. D. (1968). *Illinois Test of Psycholinguistic Abilities—Revised*. Chicago: University of Illinois Press.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University.
- Leonard, L. B., & Ritterman, S. I. (1971). Articulation of /s/ as a function of cluster and word frequency of occurrence. *Journal of Speech and Hearing Research*, 14, 476–485.
- Lewis, M. B. (1999a). Age of acquisition in face categorisation: Is there an instance-based account? *Cognition*, 71, B23–B39.
- Lewis, M. B. (1999b). Are age-of-acquisition effects cumulative-frequency effects in disguise? A reply to Moore, Valentine and Turner (1999). *Cognition*, 72, 311–316.
- Logan, J. S. (1992). *A computational analysis of young children's lexicons* (Tech. Rep. No. 8). Bloomington, IN: Indiana University, Speech Research Laboratory.
- Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, 39, 155–158.

- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1–36.
- Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122–147). Cambridge, MA: MIT Press.
- Mack, M., & Lieberman, P. (1985). Acoustic analysis of words produced by a child from 46 to 149 weeks. *Journal of Child Language, 12*, 527–550.
- Macken, M. A., & Barton, D. (1980). A longitudinal study of the voicing contrast in American English word-initial stops, as measured by voice onset time. *Journal of Child Language, 7*, 41–74.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*, 202–227.
- Marascuilo, L. A., & Serlin, R. C. (1988). *Statistical methods for the social and behavioral sciences*. New York: W. H. Freeman.
- McReynolds, L. V., & Kearns, K. P. (1983). *Single-subject experimental designs in communicative disorders*. Baltimore: University Park Press.
- Metsala, J. L. (1997a). An examination of word frequency and neighborhood density in the development of spoken word recognition. *Memory & Cognition, 25*, 47–56.
- Metsala, J. L. (1997b). Spoken word recognition in reading disabled children. *Journal of Educational Psychology, 89*, 159–169.
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Hillsdale, NJ: Erlbaum.
- Monaghan, J., & Ellis, A. W. (2002). Age of acquisition and the completeness of phonological representations. *Reading and Writing, 15*, 759–788.
- Moore, V., Valentine, T., & Turner, J. (1999). Age-of-acquisition and cumulative frequency have independent effects. *Cognition, 72*, 305–309; discussion 311–316.
- Moore, W. H., Burke, J., & Adams, C. (1976). The effects of stimulability on the articulation of /s/ relative to cluster and word frequency of occurrence. *Journal of Speech and Hearing Research, 19*, 458–466.
- Morris, P. E. (1981). Age of acquisition, imagery, recall, and the limitations of multiple-regression analysis. *Memory & Cognition, 9*, 277–282.
- Morrisette, M. L., & Gierut, J. A. (2002). Lexical organization and phonological change in treatment. *Journal of Speech, Language, and Hearing Research, 45*, 143–159.
- Morrisette, M. L., McAllister, K., & Gierut, J. A. (2009, November). *Vocabulary size and phonological complexity*. Poster presented at the annual meeting of the American Speech–Language–Hearing Association, New Orleans, LA.
- Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 3*, 528–559.
- Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 116–133.
- Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition, 20*, 705–714.
- Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47*, 1048–1058.
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*, 273–281.
- Olive, M. L., & Smith, B. W. (2005). Effect size calculations and single subject designs. *Educational Psychology, 25*, 313–324.
- Olswang, L. B. (1998). Treatment efficacy research. In C. M. Frattali (Ed.), *Measuring outcomes in speech–language pathology* (pp. 134–150). New York: Thieme.
- Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition, 38*, 43–102.

- Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders*, 52, 271–277.
- Roid, G. H., & Miller, L. J. (1997). *Leiter International Performance Scale—Revised*. Chicago: Stoelting.
- Schlosser, R. W. (2005). Meta-analysis of single-subject research: How should it be done? *International Journal of Language and Communication Disorders*, 40, 375–378.
- Shriberg, L. D., Kwiatkowski, J., & Gruber, F. A. (1994). Developmental phonological disorders II: Short-term speech–sound normalization. *Journal of Speech and Hearing Research*, 37, 1127–1150.
- Smith, B. L., McGregor, K. K., & Demille, D. (2006). Phonological development in lexically precocious 2-year-olds. *Applied Psycholinguistics*, 27, 355–375.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174–215.
- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44, 1321–1337.
- Storkel, H. L. (2004). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research*, 47, 1454–1468.
- Storkel, H. L., & Morrisette, M. L. (2002). The lexicon and phonology: Interactions in language acquisition. *Language, Speech, and Hearing Services in Schools*, 33, 24–37.
- Tainturier, M. J., Johnson, E., Tamminen, J., & Thierry, G. (2003). An ERP investigation of age-of-acquisition effects in spoken word recognition. *Brain and Language*, 87, 15–16.
- Thompson, C. K. (2007). Complexity in language learning and treatment. *American Journal of Speech–Language Pathology*, 16, 3–5.
- Tyler, A. A., & Edwards, M. L. (1993). Lexical acquisition and acquisition of initial voiceless stops. *Journal of Child Language*, 20, 253–273.
- Vitevitch, M. S. (1997). Tongue twisters reveal neighborhood density effects in speech production. *Research on Spoken Language Processing*, 21, 1–8.
- Walley, A. C. (1993). The role of vocabulary development in children’s spoken word recognition and segmentation ability. *Developmental Review*, 13, 286–350.
- Walley, A. C., & Metsala, J. L. (1992). Young children’s age-of-acquisition estimates for spoken words. *Memory & Cognition*, 20, 171–182.
- Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing*, 16, 5–20.
- Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition*. Cambridge, MA: MIT Press.
- Wiig, E. H., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals—Preschool*. San Antonio, TX: Harcourt Brace.
- Williams, K. T. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Service.