Research Article

Do the Hard Things First: A Randomized Controlled Trial Testing the Effects of Exemplar Selection on Generalization Following Therapy for Grammatical Morphology

Amanda Jean Owen Van Horne,^a Marc Fey,^b and Maura Curran^a

Purpose: Complexity-based approaches to treatment have been gaining popularity in domains such as phonology and aphasia but have not yet been tested in child morphological acquisition. In this study, we examined whether beginning treatment with easier-to-inflect (easy first) or harder-to-inflect (hard first) verbs led to greater progress in the production of regular past-tense *-ed* by children with developmental language disorder.

Method: Eighteen children with developmental language disorder (ages 4–10) participated in a randomized controlled trial (easy first, $N = 10$, hard first, $N = 8$). Verbs were selected on the basis of frequency, phonological complexity, and

telicity (i.e., the completedness of the event). Progress was measured by the duration of therapy, number of verb lists trained to criterion, and pre/post gains in accuracy for trained and untrained verbs on structured probes.

Results: The hard-first group made greater gains in accuracy on both trained and untrained verbs but did not have fewer therapy visits or train to criterion on more verb lists than the easy-first group. Treatment fidelity, average recasts per session, and verbs learned did not differ across conditions.

Conclusion: When targeting grammatical morphemes, it may be most efficient for clinicians to select harder rather than easier exemplars of the target.

Children with developmental language disorder (DLD; formerly known as specific language impairment; Bishop, Snowling, Thompson, Greenhalgh, & CATALISE-2 consortium, 2016) show a protracted language learning trajectory, and as such, they often require language intervention. In the preschool years, grammatical morphemes are frequent targets of language intervention for these children. Clinicians often take a developmental approach to teaching grammatical morphemes. They assume that beginning with easier, earlier acquired targets will lead to greater or more rapid learning than beginning with those that are harder or later acquired (e.g., Crystal, 1985; Weiler, 2013). This assumption is consistent with the theoretical construct of the zone of proximal development (Vygotsky, 1978) and with more recent theories positing

that the input children hear shapes the acquisition process. This, in turn, shapes their productions and, thus, the input for the next generation of learners (e.g., the production–distribution–comprehension hypothesis; MacDonald, 2013).

In this article, we consider an alternative to the developmental approach to target selection. The alternative is built on the hypothesis that targets that best illustrate relevant linguistic contrasts are key to successful language intervention. According to this hypothesis, instead of selecting therapy targets on the basis of their progression within a developmental series (e.g., simple sentences before complex sentences), we should select them on the basis of underlying linguistic organization. This holds even if this means that the child's initial attempts at using these targets will be less accurate than would be the case with a developmental approach. The specific problem space for the study was regular past-tense *-ed* for children with DLD. Some verbs are more or less likely to be correctly inflected with regular past tense by speakers of the language. It is well known that the phonological form of the verb stem influences inflection accuracy (e.g., Tomas, Demuth, Smith-Lock, & Petocz, 2015), but stem frequency, inflected form frequency, and event semantics also influence accuracy. Within the area of event

^aUniversity of Iowa, Iowa City

^bUniversity of Kansas, Lawrence

Correspondence to Amanda Jean Owen Van Horne, who is now at the University of Delaware, Newark: ajovh@udel.edu

Editor-in-Chief: Sean Redmond

Editor: Lisa Archibald

Received January 1, 2017

Revision received April 12, 2017

Accepted May 8, 2017

https://doi.org/10.1044/2017_JSLHR-L-17-0001

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

semantics, *telicity* refers to the completed nature of the event. For example, verbs that are rated as high in telicity reflect completed event semantics (e.g., *jumped*, *tripped*, *spilled*). Both children and adults tend to use these verbs and accurately inflect them in past-tense contexts. In contrast, verbs that are rated as low in telicity (atelic) typically depict actions and events that are not complete (e.g., *walk*, *run*, *cry*). These atelic verbs are more likely than telic verbs to be used and inflected in the present tense or with progressive aspect (e.g., *walking*, *crying*; Owen Van Horne & Green Fager, 2015; Wulff, Ellis, Römer, Bardovi-Harlig, & Leblanc, 2009).

Thus, past tense is a better fit and appears to be easier with verbs high in telicity, whereas progressive and present-tense contexts may be easier on verbs low in telicity. In this study, we queried if beginning treatment on past-tense verbs that are more difficult (i.e., less telic or hard first) leads to greater progress in therapy, as measured by faster progress in training to criterion, greater gains in accuracy on trained words, and greater generalization to untrained words than beginning treatment with verbs in the less difficult category (i.e., more telic or easy first). The results have implications for understanding how form and meaning interact in grammar learning and for understanding the role of complexity in language learning.

Grammatical Morphemes as Categories of Meaning

In her early work on construction grammar, Goldberg (1995) argues that syntactic frames (e.g., *He verb-ed the noun the noun*) carry meaning separate from the lexical items that appear in that frame (e.g., *give*). This separate meaning can be readily detected in two ways. First, when novel lexical items are inserted in the frame, listeners can still accurately gloss a meaning. For instance, a sentence such as *she glorped the ziv the wug* is interpretable as transferring an object to a new location even though the listener doesn't know any content words in the sentence because he or she has access to similar constructions using real words (e.g., *she gave the girl the book*). Second, the unique meaning of the construction can be detected when words that typically occur in other sentence frames are inserted into the construction and take on a new meaning. For instance, inserting the intransitive verb *sneeze* into a ditransitive frame, *the elephant sneezed the canary across the room*, allows the listener to interpret *sneeze* as a manner of motion in addition to a bodily function. Goldberg further argues that all aspects of language, from words and idioms to grammatical morphemes and syntactic frames, carry meaning and that the differences between the lexicon and grammar have to do with how fixed the elements of the construction are rather than with the need to link form and meaning. That is, a word (e.g., *cat*) and an idiom (e.g., *kicked the bucket*) are both highly fixed; changes in the form alter the meaning (e.g., *she was kicking the bucket* or *she kicked the door* do not have the same meaning as *she kicked the bucket*, just as changing the phonemes in *cat* leads to new meanings, such as *hat*, *cut*, *car*). In contrast to individual words

and idioms whose meaning depends on relatively fixed forms, grammatical constructions maintain their meaning in the face of change, as illustrated by the use of novel words and unusual verbs in the transfer of location construction above.

Goldberg's (1995) claims lay the foundation for Ambridge, Pine, and Rowland's (2011) FIT hypothesis, which stands for frequency, item, template. In the FIT hypothesis, Ambridge et al. (2011) claim that the production of a word in a grammatical construction is determined probabilistically on the basis of the frequency of the word itself, the frequency of the item in the construction or template, and the semantic relevance or FIT of the word for the situation. Thus, they argue that both semantic features of the specific words being used and probabilistic exposure to lexical items in predictable templates or constructions guide acquisition of morphology and syntactic frames (Ambridge & Lieven, 2011; Ambridge, Pine, Rowland, Chang, & Bidgood, 2013). Although the FIT hypothesis has been most well developed for how verbs fit into syntactic frames, it is also applicable to morphology. All verbs can combine with all tense and agreement morphemes, but some verbs probabilistically appear more or less frequently with particular tense and agreement morphemes, meaning that both the frequency of the verb itself and the frequency of the verb paired with the morpheme matter. That is, it is rare to hear sentences such as *I am wanting an ice cream* or *she is knowing her friend*, even though it is possible, because these verbs have a poor fit with the present progressive morphological template. This probabilistic pattern of verb and morpheme use is hypothesized to inform both production and acquisition.

Li and Shirai (2000) examined the acquisition and use of tense and aspect in early language development and found that English-speaking children initially produce regular morphology like past-tense *-ed*, third person singular *-s*, and the progressive participle *-ing* with words that have lexical semantics closely aligned with the respective tense marker. In particular, they examined the role of lexical aspect, the time information inherent to the event described by the verb. Thus, *-ed* is most commonly heard and produced on verbs that are telic, such as *closed* and *dropped*, as opposed to atelic verbs, such as *walked* and *rolled*. Third person singular *-s* is most commonly heard and produced on stative verbs that denote ongoing states of being, such as *knows* and *wants*, and the progressive participle *-ing* is most commonly heard and produced on atelic, activity verbs that denote ongoing actions, such as *wiggling* or *coloring*. Li and Shirai speculate that this early alignment of tense and lexical aspect helps children discover the meaning of tense-related morphemes.

Thus, much like there are good (*robin/bluebird*) and poor (*ostrich/penguin*) examples of a semantic category such as *bird*, there are also good (*dropped*) and poor (*wiggled*) examples of verbs whose meaning aligns with a particular verb marker such as past-tense *-ed*. Although it is easiest to think of *dropped/wiggled* and *robin/ostrich* as good/poor exemplars of a category, typicality is confounded with

frequency and distributional information within the category. *Robin*, for example, is a good exemplar of the category *bird* because it occurs frequently and shares features with other examples (e.g., *bluebird/cardinal*; Borovsky & Elman, 2006; Posner & Keele, 1968). In a similar manner, both frequency and semantics influence morphology production (Arnon & Clark, 2011; Bloom, Lifter, & Hafitz, 1980; Oetting & Horohov, 1997). Children tend to hear verbs that are semantically related to the inflection more often with that particular inflection, making it difficult to separate verb + inflection frequency and verb semantics. In fact, the same factors are critical components in accurate production by English language learners with the most frequent verb + morpheme combinations produced most accurately and these frequent combinations reflecting high alignment between tense and aspect information (Wulff et al., 2009).

The factors that seem aligned with early inflection use are not necessarily those that promote learning, however. Using a computational model, Li and Shirai (2000) also demonstrated that children begin to differentiate the morphemes from the verbs they are associated with when they hear morphemes with less well-aligned verbs. So hearing a stative verb, such as *wanted*, with a past-tense ending; a telic verb, such as *dropping*, with an *-ing* ending; or an activity verb, such as *colors*, with a present/habitual third person singular *-s* marker might indicate that the morpheme adds a unique meaning to the verb and lead to the discovery of a broader category of tense. It is critical that the input that triggers reorganization of what the morphemes contribute to verb meaning is not the same input that promotes early accurate productions by children. On the basis of the computational model, Li and Shirai hypothesized that one way to accelerate morpheme learning would be to expose children to unusual verb + morpheme pairs, such as those described previously, to illustrate the boundaries of how the morphemes are used and how they modulate the meaning of verbs. To our knowledge, this hypothesis has not yet been tested with either children or adult learners.

Phonological Complexity

One well-documented area of difficulty for children with DLD is producing phonologically complex forms (Leonard, Davis, & Deevy, 2007; Marshall & van der Lely, 2007). Although some researchers have documented phonological complexity on the basis of the number of phonemes in a cluster when a verb is inflected (e.g., /mpt/ in *jumped* vs. /nd/ in *stunned*; Marshall & van der Lely, 2007) and others have based it on the phonotactic probability of the inflected forms (Leonard et al., 2007), all researchers have shown this complexity to be extraordinarily challenging for children with DLD. In addition, the phonological patterns associated with some stems may lead to confusion with regard to whether or not past tense is required. This holds for a variety of populations (e.g., aphasia; Rimikis & Buchwald, 2015) and not just children (Marchman, Wulfeck, & Ellis Weismer, 1999). In

particular, computational models have shown that stems ending with *-t* and *-d* and taking the syllabic allomorph (e.g., *start + ed*, *waste + ed*) as the inflection are much less likely to be inflected accurately. As with the confound between frequency and semantics described previously, similar confounds exist within the phonological pattern. It is unclear whether problems with the syllabic allomorph are due to the fact that the brain registers *-t* and *-d* as potential evidence that the verb is already inflected (Marchman et al., 1999) or because the syllabic allomorph is much less frequent (Tomas et al., 2015). Nonetheless, the majority of the word-final clusters formed to produce the regular past-tense form are unattested or low in frequency as clusters in monomorphemic words, and this cluster complexity can interfere with accurate suffix productions.

Some of the earliest work using complexity-based treatment approaches arose from Gierut's lab in the early 1990s. She proposed that maximal, rather than minimal, opposition would lead to the greatest generalization because this would introduce the greatest number of features into a child's repertoire (Gierut, 1989). She demonstrated that treating complex three-element clusters (e.g., *spl-*) led to generalization to the subcomponents, including singleton fricatives (e.g., *s-*) and untreated two-element clusters (e.g., *pl-* or *sp-*; Gierut & Champion, 2001). Given the contribution of phonological complexity to the challenges of producing the past-tense inflection, one might wonder if choosing low-frequency or unusual phonological combinations might also lead to gains in phonemic production. Gierut's work has been challenged extensively (Rvachew & Nowak, 2001; Rvachew, Rafaat, & Martin, 1999), with some researchers arguing that beginning with the most stimutable phonemes first is more pleasant for the child and leads to greater gains than when more complex sounds are targeted in therapy. A lab not affiliated with either position recently showed no difference in duration of treatment between the two approaches (Dodd et al., 2008).

A complexity-based approach to inflection might assume that learning to inflect in cases in which the articulatory or cognitive challenges posed by the morphophonological form are high would lead to acquisition of the simpler, more frequent, or more transparent cases. The contrasting perspective would be a developmental approach that assumes that one should ensure that the child is capable of producing the target inflection before treating it in therapy—that is, that past tense should only be targeted once word-final *-t* and *-d* are established as singletons and in monomorphemic clusters (e.g., *tent*, *tilt*). If it is not possible to elicit clusters, then exemplars should be selected that are within the child's phonological repertoire (e.g., *played*). Low-frequency clusters, such as the *-mpt* in *jumped*, should be treated only at the conclusion of therapy.

Combined Effects of Morphophonology and Semantics

Prior work done by our labs has shown that morphophonology and telicity combine to influence accuracy of

productions by typically developing children and children with DLD in the moment (Johnson & Fey, 2006; Owen Van Horne & Green Fager, 2015). Using elicited production, Johnson and Fey (2006) demonstrated that both phonological complexity and telicity independently contributed to changes in accuracy of past-tense production for typically developing 2-year-olds and for a single child with DLD. Thus, the best performance was observed in verbs that depicted completed events and a stem that ended in a non-obstruent consonant. The worst performance was observed in verbs that had multiple nonfacilitative characteristics, such as verbs that represented incomplete events and stems ending in obstruent consonants. Graded performance was observed for those verbs having a mix of facilitative and nonfacilitative characteristics. Inspired by this work, Owen Van Horne and Green Fager (2015) extended and replicated their findings with an older cohort of children, a larger DLD group, and a wider range of verbs. Thus, it is possible to distinguish “hard” verbs—verbs that depict atelic events, that are rarely inflected with the past tense, and that end in alveolars or obstruents—from “easy” verbs that depict telic events, that are commonly inflected with the past tense, and that end in sonorants. These verbs can be placed along a continuum of difficulty. Although we know that these complexity factors influence accuracy in the moment for a variety of populations, including children with DLD (Johnson & Fey, 2006; Owen Van Horne & Green Fager, 2015), adult speakers of English as a second language (Wulff et al., 2009), and child bilinguals (Blom & Paradis, 2013; Rispens & De Bree, 2014), we do not know how exposure to verbs at different points on the continuum influences learning.

Questions

In this study, we asked if exemplar selection influences the efficacy of treatment for past tense in children with DLD. We addressed this question by selecting verbs for treatment on the basis of whether they were “easy” or “hard” to inflect accurately. To make this determination, we used a composite rating comprising lexical aspect ratings, frequency of the inflected form in the Child Language Data Exchange System (MacWhinney, 2000), and the morphophonological properties of the stem + inflection (see Owen Van Horne & Green Fager, 2015, for more details). We assessed treatment efficacy by answering three questions:

1. Do the children assigned to the easy-first or hard-first condition make faster progress in therapy?
2. Do the children assigned to the easy-first or hard-first condition make greater gains in accuracy from pretest to posttest on trained verbs?
3. Do the children assigned to the easy-first or hard-first condition make greater gains in accuracy from pretest to posttest on untrained verbs held out to test for generalization?

Method

This early efficacy intervention study used a randomized controlled trial to test our hypothesis that certain types of verbs would lead to greater generalization of regular past-tense *-ed* to new verbs. One cohort proceeded from easy verbs to hard verbs, and the other proceeded from hard to easy. The primary dependent variables were (a) the number of visits necessary to progress through all six lists of verbs up to a maximum of 36 visits and (b) differences between pretest and posttest accuracy for target and generalization verbs. Reporting follows the Consolidated Standards of Reporting Trials (CONSORT) 2010 guidelines (Schulz, Altman, & Moher, 2010).

Research procedures were approved by the University of Iowa Human Subjects Board and completed in compliance with the ethics guidelines outlined in the Belmont Report (National Commission for the Protection of Human Subjects, 1978). Procedures were carried out between August 2013 and January 2017.

Participants

A priori power analyses on the basis of other treatment studies suggested that a sample size of 20 would be required. Participant recruitment focused on children ages 4–9 who were receiving speech-language services across the state of Iowa. In addition to contacting speech-language pathologists directly, we contacted principals, schools, tutoring centers, and private practice clinics and placed advertisements in local periodicals. Children’s parents contacted us via phone or email, and then an initial phone screening was followed by one or more in-person screening visits.

Children were initially identified as meeting the inclusionary and exclusionary characteristics for DLD, and then their use of past-tense *-ed* and final consonants was documented. Enrollment focused on children’s ability to use target structures and meet the DLD criteria rather than age. Children were identified as having a disorder in language development if they obtained a standard score of 95 or less on the Structured Photographic Expressive Language Test—Third Edition (Dawson, Stout, & Eyer, 2003), following recommendations of Perona, Plante, & Vance, (2005), a highly sensitive and specific test that is normed for children ages 4–9. To confirm normal nonverbal intelligence, all children obtained a standard score above 80 on the Kaufman Brief Intelligence Test—Second Edition, matrices subtest (Kaufman & Kaufman, 2004). Children also passed a hearing screening (American Speech-Language-Hearing Association, 1997) and, with the exception of a diagnosis of attention deficit hyperactivity disorder, had no additional medical, psychiatric, or cognitive disorders. In addition to qualifying as having DLD, children were also required to (a) produce word-final /t/ and /d/ with at least 80% accuracy; (b) combine subjects and verbs (Hassink & Leonard, 2010); (c) use regular past-tense *-ed* in a 60-item elicitation task adapted from Redmond and Rice (2001) less than 40% of the time; and (d) produce the target verb,

even if it was uninflected, for at least 30 of the 60 items in the elicitation task. Monomorphemic production of /t/ and /d/ was usually nonimitative (Examiner [E]: *a car drives on ...* Child [C]: *the road*) but could be elicited via imitation if needed (E: *a car drives on ...* C: *wheels*. E: *say road*). In addition, the Peabody Picture Vocabulary Test–III (PPVT–III; Dunn & Dunn, 1997) and Expressive Vocabulary Test (Williams, 1997) were administered to provide descriptive information about participants' vocabulary abilities.

Table 1 reports the initial test scores for children in the two experimental groups and whether or not they were receiving speech-language services at the time of intervention. Groups did not differ on any demographic variables, with *p* values ranging from $t(16) = 1.52, p = .14$ (accuracy on *-t* and *-d*) to $t(16) = 0.31, p = .76$ (maternal education level). All children were speakers of standard American English on the basis of clinical observation and had less than 20% exposure to another language per parent report. It is worth noting that two children were exposed to a language other than English when staying with a noncustodial parent (approximately one weekend every other week). One additional child was exposed to Korean before 1 year of age but was exposed to only English after adoption. Participants primarily identified themselves as White ($n = 12$) with the remaining participants identifying as another race or ethnicity (Hispanic: $n = 1$, Black: $n = 1$, Asian: $n = 1$, multiple races: $n = 2$) or not reporting race ($n = 1$) on the case history form.

The enrollment process is illustrated in Figure 1. Phone and email screenings were not tracked. One hundred twenty-nine children were screened in person for participation. Twenty-seven children qualified for the study, and 23 agreed to enroll. Two children, one from each condition, withdrew after completing one or more treatment visits; both cited scheduling challenges. Three children's data were excluded after completing the study. Information about

the excluded children appears in Table 2. This resulted in 18 children with usable data completing the study.

Qualifying children were randomly assigned to treatment conditions on the basis of the order in which they enrolled. A sequence of assignments was generated by the first author using a random number generator, with odd numbers assigned to the easy-first condition and even numbers assigned to the hard-first condition. Assessors had access to the randomization order during testing. Children and caregivers were blind to the assigned condition. Treatment providers were blind to the purpose of the study and the reason why they were working on a particular set of words. On occasion, to accommodate family scheduling needs and student holidays, the lab manager or a doctoral student with speech-language pathology credentials filled in, and these individuals were not blind to treatment condition.

Treatment

Stimuli

Verbs were selected for use in therapy and probes on the basis of their rankings by Owen Van Horne and Green Fager (2015). Owen Van Horne and Green Fager rated 60 regular past-tense verbs on a continuum with regard to predicted production accuracy using a mixed model logistic regression. The standardized beta values (allowing comparison across predictors on different scales) indicated that events that were rated as highly telic (e.g., *close*, *drop*) and verbs frequently heard with an *-ed* inflection led to greater accuracy with *-ed* (e.g., *play*, *remember*). Verbs heard frequently in the stem form (e.g., *fish*) and verbs that ended in alveolars or obstruents (e.g., *rest*, *drop*) had decreased accuracy. It was surprising that each of these factors made approximately equal contributions to the model. Thus, verbs that were highly telic (i.e., completed in nature, such as *drop*), commonly inflected with past tense, rarely heard in the bare stem form, and ended in nonobstruent and nonalveolar consonants were rated as "easy" verbs. Verbs that were atelic (i.e., ongoing in nature, such as *wiggle*), rarely inflected with past tense, frequently heard in the bare stem form, and ended in obstruent and alveolar consonants were rated as "hard" verbs. The continuum of verbs was constructed with these semantic and phonological factors in mind.

Verbs were divided into two sets matched for difficulty: a treatment set and a generalization probe set. The treatment set was sorted into six lists of five verbs each, such that the lists were ordered from easiest (List 1) to hardest (List 6). The generalization set was sorted into six lists of five verbs each, such that each list was of approximately equal difficulty (see Appendix A). Children who were randomized to the easy-first condition began with List 1 and progressed through to List 6. Children who were randomized to the hard-first condition began with List 6 and progressed in the reverse order to List 1.

Treatment Procedures

In accord with current best practice, we explicitly used a mix of drill-based, modeling, and more naturalistic

Table 1. Demographic characteristics of children enrolled in the randomized controlled trial.

Characteristic	Easy first ^a	Hard first	<i>p</i>
<i>N</i> (girls, boys)	10 (3, 7)	8 (3, 5)	
<i>n</i> in speech/language therapy	7	6	
Age in months	63.1 (23.02)	72.75 (18.48)	.35
Maternal education in years	16.2 (1.87)	15.88 (2.64)	.76
SPELT-3 SS	80.88 (8.23)	78.00 (11.25)	.55
KBIT-2 SS	102.2 (11.47)	99.62 (6.69)	.58
PPVT-III SS	102.6 (9.31)	99.5 (8.88)	.48
EVT SS	97.5 (9.68)	92.37 (15.05)	.39
Percent accuracy <i>-t/-d</i>	98 (4.22)	93.75 (7.44)	.14

Note. SPELT-3 = Structured Photographic Expressive Language Test—Third Edition; SS = standard score; KBIT-2 = Kaufman Brief Intelligence Test—Second Edition; PPVT-III = Peabody Picture Vocabulary Test—III; EVT = Expressive Vocabulary Test.

^aOne child in the easy-first condition was 10;0 (years;months) old when enrolled and therefore technically outside of the norms of the SPELT-3. Using 9;6–9;11 norms, the child scored below the first percentile. His score is not included in the average here.

Figure 1. Flowchart illustrating participant flow through enrollment procedures and treatment protocol. DLD = developmental language disorder.

approaches to maximize effectiveness for a variety of age ranges and learning styles (Eisenberg, 2013). Recasting is an effective intervention strategy (Cleave, Becker, Curran, Owen Van Horne, & Fey, 2015) shown to lead to gains of 0.7–1.0 *SD* at rates of approximately 0.6–1.0 recasts/min. Recasting arguably promotes generalization to non-therapeutic contexts (Camarata, Nelson, & Camarata, 1994; Mohammadzaheri, Koegel, Rezaee, & Rafiee, 2014).

At each treatment visit, the child learned to inflect the verbs from one of the lists of treatment verbs. Treatment visits began with a 10-item sentence imitation task that included each verb presented twice: once sentence medially and once sentence finally (Dalal & Loeb, 2005). This task served both to provide a drill-based context for the child to

learn the verbs and as a means to assess progress from session to session. For each sentence imitation item, clinicians followed a prompting hierarchy, described in Appendix B, to cue the child for incorrect responses (two additional exposures per sentence) or provide positive feedback in the case of correct responses (one additional exposure per sentence). Including the exposures in the sentences to be imitated and the response hierarchy meant that children heard 20–30 total examples of the past-tense *-ed* (four to six examples per verb). Children who scored lower than 80% correct on the uncued sentence imitation continued with a treatment visit (described in the section on treatment visit protocols); children who scored 80% or more correct continued with a probe visit (described in the section on outcome

Table 2. Demographic characteristics of three children who completed treatment but were excluded from analyses.

Demographic variable	2298	1994	2074
Condition	EF	EF	HF
Gender	F	M	M
Enrolled in therapy?	N	N	Y
Age in months	78	57	50
SPELT-3 SS	76	80	58
KBIT-2 SS	83	88	82
PPVT-III SS	91	58	87
EVT SS	85	92	91
Percentage accuracy <i>-t/-d</i>	100	90	100
Reason for exclusion	Pretest probes miscored, past tense too high	Epilepsy diagnosis 6 months poststudy	Psychiatric disorder; significant behavior concerns during treatment

Note. EF = easy first; HF = hard first; SPELT-3 = Structured Photographic Expressive Language Test—Third Edition; SS = standard score; KBIT-2 = Kaufman Brief Intelligence Test—Second Edition; PPVT-III = Peabody Picture Vocabulary Test—III; EVT = Expressive Vocabulary Test.

measures). The flow through the treatment protocol is illustrated in Figure 2, a flow chart that was available to intervention providers to support their decision making.

Treatment visit protocols. Following the sentence imitation task, children heard each verb inflected with the past tense five times via observational modeling (Leonard, 1975) for a total of 25 exposures. Next, syntax stories alternated with focused stimulation for the remainder of the therapy visit. A syntax story is a story loaded with the particular grammatical construction being targeted. Children heard a single story and then engaged in play to retell the story for 5–10 min. The clinician then read another syntax story and engaged in play to retell the story for another 5–10 min.

Each syntax story focused on between one and three verbs and included at least three exposures to each verb. These stories both provided models of the verb and served as a platform for promoting use of the target verbs for recasting. Verbs were divided among several shorter stories rather than a single long story because piloting showed that this promoted participation. During play, the examiner modeled and recasted the target verbs in past tense. For instance if the child said, “Frog jump,” the examiner responded, “The frog jumped.” Play and stories alternated for 25 min or until all five verbs had been recast three to five times. After they achieved the target of five recasts for a verb, the examiner was instructed to move to another verb. This resulted in approximately 65–85 models or recasts

Figure 2. Flowchart of participant progression through study activities.

per visit and a rate of 0.6 to 1.0 recasts/min during the focused stimulation portion of the treatment sessions. A sample treatment visit script is available in Appendix B. For the oldest children or for children who were focused on a single list for a long time, a variety of adaptations were allowed to maintain child engagement. For example, the examiner might draw with the child instead of using toys. Other toys might be added to the usual toy set, or new exemplars of the original objects might be introduced to keep the children engaged.

Treatment fidelity. Because accommodating child schedules was a priority and children were seen in multiple locations across the state, 16 different people provided intervention. All providers had a background in speech-language pathology or elementary education, with credentials ranging from being enrolled in an undergraduate communication sciences and disorders program to being a fully certified speech-language pathologist. Because of this, treatment fidelity was maintained in a variety of ways. Initial training included having providers read a treatment manual and watch a training video containing clips from therapy sessions, then talk through procedures with the first author or the lab manager. Next, the provider observed a live treatment session. Following this session, the lab manager or the first author provided as many coaching visits as needed for both the trainer and the provider to feel confident that the provider could carry out intervention independently. Most providers observed one session live and then had two to three coaching sessions, but some asked to observe more sessions or asked for more coaching. Providers were given a written script to use for each session that included space to indicate responses to sentence imitation, how many observational modeling attempts were completed for each verb, the stories to read, and a tally sheet for the number of recasts completed per verb.

In addition, each visit was audio recorded. In total, 293 of the 553 completed treatment sessions were checked for fidelity. Faithfulness to the administration of the protocol was documented for 25%–100% of the sessions for each child–provider pair. These fidelity checks provided both an opportunity for formative feedback to be provided as therapy progressed and indicated how closely the provider was adhering to protocol. Feedback focused on eliciting difficult-to-elicited verbs, using more past-tense *-ed* models than bare stem models, decreasing the number of missed recast opportunities (even if the minimum number of recasts was achieved), and general behavior management strategies.

As shown in Table 3, for the vast majority of those sessions checked, sentence imitation, observational modeling, and syntax stories were completed as expected. The primary source of unwanted variability across visits was recasts. In this regard, 89% of the target verbs were recast at least three times per session. In a similar manner, 90% of visits met minimum recast goals of at least 15 recasts per visit, with many exceeding that goal. The primary reason for failing to meet recast goals was difficulty eliciting platform utterances containing certain target verbs.

Table 3. Mean (and standard deviation) of the proportion of sessions in which each aspect of the intervention was administered faithfully and the number of recasts per session for each condition.

Intervention component	Easy first	Hard first
Sentence imitation	1.00 (0.00)	0.98 (0.05)
Observational modeling	0.93 (0.10)	0.92 (0.10)
Syntax story	0.96 (0.04)	0.96 (0.06)
All five verbs recast three or more times	0.74 (0.23)	0.66 (0.32)
Recast rate	26.85 (5.70)	26.49 (9.79)

Outcome Measures

Pre-/Posttreatment Probes

Probes on the basis of Redmond and Rice (2001) were completed for all 60 possible verbs, including all verbs from the treatment and generalization sets. The probes used puppet shows and parallel structure to obligate the use of regular past-tense *-ed*. Each probe used an irregular verb in the first clause and prompted the child to produce the target regular verb in the second clause. The target verb was modeled in a bare stem form prior to elicitation. Target actions were designed to be produced briefly and noniteratively to encourage the production of regular past-tense *-ed*. Only production of the target verb was scored; substitutions of other verbs were discarded as unscorable. A sample probe item is in Appendix C.

Pretest probes were administered in conjunction with other testing, and examiners attempted to administer approximately 20 items per pretest visit. Given that standardized tests were not repeated during the posttest phase, children typically completed 30 items per day, and post-testing generally took no more than two visits. Pre-/posttesting was primarily completed by the lab manager, an individual with a bachelor's degree in an unrelated field and extensive training on the lab protocols. The lab manager was not blind to condition allocation. Scoring decisions were made online and later verified from audio recordings made at the time of testing.

Within-Treatment Probes

Progress in treatment was made by advancing through each of the six verb lists, one at a time. If a child got 80% of the sentence imitation targets correct at the beginning of a visit, a within-treatment probe was administered in lieu of a treatment visit to determine if the child could inflect the target verbs in nonimitative contexts. As with the pre-/posttest probes, within-treatment probes consisted of a sentence completion task (Redmond & Rice, 2001). The probe included 10 items: five verbs from the treatment set that was just being targeted and five nontarget verbs. To advance to the next treatment list, children had to correctly inflect four out of five of the trained verbs; if not, the next visit was necessarily a complete treatment visit regardless of performance on the sentence imitation task. This was done because piloting showed that most children lacked the stamina to complete both a probe and a treatment visit on the same day, and we wanted to avoid having children

participate in repeated testing visits with no treatment visits. This occurred in piloting when a child had achieved 80%–100% correct on sentence imitation but was not yet succeeding at the probe task.

Although all probes were audio recorded for later verification, both sentence imitation and within-treatment probes were scored by the interventionist online. Immediate decisions about whether to do a treatment visit or a probe visit were made by the interventionist in the moment. Results from probe visits were listened to again by the interventionist as needed to verify a response, and then results were recorded and advancement decisions were made in consultation with the primary investigator and lab manager.

Scoring of Probes

Probes were initially scored online, and later, a subset of probes was listened to a second time to document reliability of scoring (see the next section on reliability). The scorer initially determined if the child attempted to produce the verb that the probe item intended to elicit, produced another verb (*hop* for *jump*), or responded in such a way that the response was unscorable (i.e., unintelligible and off-topic responses). If the child attempted the target verb, then the verb was scored for accuracy of inflection. Any attempt at past tense was accepted even if it was not phonologically correct. For instance, a child who said *jumped-id* would have received credit for a past-tense production. The response was scored as wrong if the child used a bare stem (e.g., *jump*) or used a nonpast inflection (e.g., *jumps*).

Reliability

Ten percent of all probes were independently retranscribed and rescored by research assistants blind to condition and to the hypotheses of the study. Point-by-point agreement ranged from 85% to 90%.

Statistical Approach

Although Student's *t* test is robust to violations of the normal distribution and all of our comparisons were planned comparisons, the number of participants was very small, and there were potential violations of the assumption of normality. As such, we also carried out nonparametric statistics. For most comparisons, we supplemented *t* tests with the Mann–Whitney *U* test. In general, gain scores were used to account for the pretest scores while still allowing the use of comparable parametric and nonparametric tests for between-group comparisons. Hedge's *g*, an effect size measure that corrects for small samples, was computed and reported for significant effects. In all of these cases, the parametric and nonparametric statistics yielded highly similar results, and thus, only parametric results are reported.

Nonparametric tests are reported in two cases. The first is the number of visits to complete intervention. The scale for this measure is truncated because we stopped therapy after 36 visits regardless of progress made. As such, only the median test, a nonparametric

test, meets assumptions. A post hoc analysis of the correlation between age or vocabulary and gain scores was also carried out. Scatterplots of the data suggested that assumptions of linearity and homoscedasticity were violated, requiring the use of a nonparametric measure of association, Spearman's rho.

Results

Number of Visits to Complete Intervention

Recall that children participated in intervention for up to 36 visits or until the child had tested out of all verb lists. Thus, two possible metrics of duration of visit exist: the number of lists the child participated in (ranging from one to six) and the number of visits the child participated in (ranging from 12 to 36). Regardless of the measure used, there were no statistically significant differences between groups (number of lists, $\chi^2 = 0.74$, $p = .38$; number of visits, $\chi^2 < 0.00$, $p > .99$).

One could imagine that no significant differences might hide a trend. That is, children might begin slowly and then speed up (the hypothesized path in the hard-first condition) or progress steadily through the lists (the hypothesized path in the easy-first condition). A qualitative examination of progress also indicates no clear patterns in the treatment responses across the groups (see Tables 4 and 5).

Changes in Accuracy of Target Verbs

Accuracy data are reported in Table 6. We first asked if there were group differences in children's use of past-tense accuracy at pretest on the 30 target verbs. No differences were obtained, $t(16) = 0.89$, $p = .38$.

Next, we assessed gains in accuracy on the 30 target verbs. Gain scores were calculated by subtracting pretest accuracy from posttest accuracy. Gains in accuracy for target verbs were significantly greater for the hard-first group, $t(16) = 2.46$, $p = .026$. This effect was large, Hedge's $g = 1.16$.

One concern was that some children did not complete all of the training lists. In fact, only three of 10 children in easy first and four of eight children in hard first completed all of the lists. Thus, many children did not receive training on all 30 verbs. The range across children

Table 4. Number of visits the child spent on each list (List 1 = easiest verbs, List 6 = hardest verbs) for the easy-first condition.

Subject	List 1	List 2	List 3	List 4	List 5	List 6
1597	1	5	8	4	2	2
2641	5	6	8	4	2	6
2240	6	8	11	2	2	3
2827	9	6	20	1		
2500	12	23	1			
2645	5	31				
2916	30	6				
2659	36					
2471	36					
2485	36					

Table 5. Number of visits the children spent on each list (List 6 = hardest verbs, List 1 = easiest verbs) for the hard-first condition.

Subject	List 6	List 5	List 4	List 3	List 2	List 1
2921	2	2	2	2	2	2
2552	3	2	4	3	4	1
2182	3	6	2	4	2	2
2115	5	7	2	2	3	2
2587	4	11	9	12		
2634	7	21	4	4		
2320	16	4	14	2		
2800	36					

was five to 30 trained verbs (one to six lists). We recalculated the gains in accuracy, only including those verbs from lists that the child had actually been trained on (e.g., Lists 3, 4, 5, and 6 for Subject 2587) and found no group differences, $t(16) = 1.73$, $p = .10$. It should be noted that this is an especially stringent test of gains because the hard-first group would thus have a greater proportion of “hard” test items while the easy-first group would have a greater proportion of “easy” test items. It is perhaps not surprising then that the comparison was nonsignificant even though visual inspection of the data suggests that the hard-first group made greater gains.

Changes in Accuracy of Generalization Verbs

As with the analysis above, first, we verified that there were no group differences on the 30 generalization verbs at pretest, $t(16) = 0.09$, $p = .93$. Next, we assessed gains on the 30 generalization verbs. Gains in accuracy for generalization verbs were significantly greater for the hard-first group as compared with the easy-first group, $t(16) = 4.06$, $p = .0009$, $g = 1.93$.

Because some children were never exposed to certain training verb lists, these untrained verbs could also be considered candidates for generalization (e.g., Lists 1 and 2 for Subject 2587). The hard-first group made greater gains

than the easy-first group on all untreated verbs, $t(16) = 3.70$, $p = .002$, $g = 1.76$.

Performance of Individual Children

Given the wide range of ages and vocabulary skills observed in the children, it is possible that either age or vocabulary was the true reason for differences in conditions. Figures 3 and 4 illustrate the observed gains in accuracy against age and vocabulary, respectively. Correlations for age and vocabulary were computed without regard for group assignment. Raw PPVT-III score was moderately correlated with treatment gains, $\rho = .47$, $p = .048$, but age was not significant, $\rho = .38$, $p = .12$.

It is also notable that four children never advanced beyond the first list. Examination of test scores and other information about these children suggests that they were among the most severe cases, scoring among the lowest six children for Structured Photographic Expressive Language Test–Third Edition percentiles and pretest scores on the past-tense probes. However, they were not the youngest, nor did they have the lowest scores on other measures. It is surprising that lack of progress on the training lists did not preclude gains in accuracy at posttest. Two of the children (one each in hard first and easy first) made gains of approximately 0.30 at posttest while the other two (both in easy first) made only minimal gains or regressed in accuracy.

Alternative Explanations

One possibility is that children made greater gains in producing tense marking on the probes because they learned more verbs in one condition than the other and were able to provide more scorable responses. Given that the hard verbs were less familiar than the easy verbs, perhaps this had more to do with a growing vocabulary than with exposure to tense marking. Recall that we only scored responses in which the child attempted the target verbs in the

Table 6. Mean (and standard deviation) pretest and posttest scores and number of verbs attempted for target and generalization verbs.

Scoring method	Condition	Pretest				Posttest			
		Target		Generalization		Target		Generalization	
		Number of verbs	Proportion correct	Number of verbs	Proportion correct	Number of verbs	Proportion correct	Number of verbs	Proportion correct
Planned ^a	Easy first	24.40 (3.20)	0.17 (0.13)	23.80 (3.58)	0.25 (0.18)	27.80 (2.20)	0.40 (0.29)	26.70 (1.95)	0.38 (0.26)
	Hard first	24.50 (4.44)	0.22 (0.12)	24.25 (4.62)	0.25 (0.15)	28.00 (2.20)	0.68 (0.17)	27.88 (2.03)	0.63 (0.17)
Actual ^b	Easy first	0.18 (0.16)	12.80 (7.66)	0.22 (0.16)	35.40 (11.09)	15.50 (10.38)	0.46 (0.33)	39.00 (9.01)	0.35 (0.25)
	Hard first	0.20 (0.13)	19.38 (8.58)	0.26 (0.14)	29.38 (6.74)	21.63 (7.91)	0.69 (0.16)	34.25 (7.69)	0.65 (0.16)

^aPlanned scoring means that all verbs pre-planned as target and generalization verbs were analyzed in those categories. ^bActual scoring means that only those verbs actually treated were analyzed as target verbs and any verbs that did not get direct treatment—because the child did not make enough progress to reach those lists—were analyzed as generalization verbs.

Figure 3. Scatterplot of gains in accuracy over the course of treatment expressed as a proportion plotted against age expressed in months.

probes. A common reason why children failed to attempt the target verb was lack of familiarity with that verb or preferring a more frequent verb with a similar meaning (e.g., *sleep* vs. *rest*). Thus, to assess changes in vocabulary due to treatment, we looked at if children were more likely

to attempt to produce a target verb during the probes regardless of whether it was inflected. No group differences were observed for the number of verbs attempted at pretest, $t(16) = 0.16, p = .87$, or at posttest, $t(16) = 0.76, p = .45$ (see Table 6).

Figure 4. Scatterplot of gains in accuracy over the course of treatment expressed as a proportion plotted against raw Peabody Picture Vocabulary Test–III (PPVT–III) scores.

It is also possible, given the variations in dosage, that some children received better or more focused stimulation than others and that this was distributed unevenly across groups. Drawing on fidelity logs, there were no differences between groups in two relevant metrics: the average number of total recasts per session regardless of verb, $t(16) = 0.09$, $p = .92$, and the number of sessions in which all five target verbs were recast at least three times each, $t(16) = 0.63$, $p = .53$ (see Table 3).

Discussion

In this study, we aimed to determine whether beginning therapy with phonologically simple, frequently inflected, telic verbs (easy first) or with phonologically complex, rarely inflected, atelic verbs (hard first) led to faster progress in therapy and/or more gains on the set of trained verbs or a set of untrained generalization verbs. Children who began therapy in the hard-first condition made greater gains on the overall verb set. Additional analyses suggest that these differences were attributable to gains made on verbs that were never trained, either because the child did not get to that point in the training or because the verbs were reserved specifically for use in testing generalization. These results could not be attributed to changes in the number of target verbs known to the child or differences in treatment administration. There were no differences in the time in therapy or the progress made on the verbs that were actually targeted during training when only those verbs are analyzed. The evidence suggests a role for vocabulary such that children with larger overall vocabulary sizes make greater gains.

This study is limited primarily by the small sample size. To ensure a fair test of the hypothesis that beginning with complex targets led to greater gains in production accuracy for untrained verbs than beginning with simple targets, stringent enrollment criteria were used, which may have led to a sample that is less representative of children typically seen in a speech-language clinic. In particular, it is likely unusual to see children who have relatively good articulation and difficulty with regular past tense. It is certain that a larger sample size or studies that replicate these findings with other morphological forms would make these findings more robust and allow analyses that would support strong recommendations for beginning intervention efforts with more difficult verb stems. A larger study would also allow us to determine if differences in vocabulary are the reason for group differences in treatment outcomes given that the hard-first group had numerically (but not statistically) higher vocabulary scores if vocabulary size moderates an otherwise meaningful group difference. As it is, an additional 10 children would be required in order to carry out interpretable covariate analyses.

It may also appear that difficulty in maintaining treatment fidelity limits the findings. However, it is worth pointing out that many components of the treatment were delivered with very high accuracy (sentence imitation, observational modeling, syntax stories, overall recast rate),

and the primary concerns had to do with whether all five target verbs were recast in every therapy session at least three times. Some children avoided or refused to say the target verbs, making it difficult for interventionists to recast the verbs. Nonetheless, this limitation worked against our hypothesis, with slightly higher, although not statistically different, recast rates occurring in the easy-first condition as compared with the hard-first condition. More variability was also documented in the recast rates for the hard-first condition, suggesting that some children found these verbs to be more challenging to use in spontaneous language. In fact, observational modeling was added to the protocol because the pilot children found the hard-first verbs to be very difficult, and clinicians felt the need to teach the verbs prior to engaging in focused stimulation. Before this approach can be deployed in a clinical setting, it will be necessary to find ways to elicit challenging verbs.

In their original model, Li and Shirai (2000) linked increasing verb vocabulary size to recognition that morphological forms add unique meanings to the verbs. In fact, they argue that it is this ever-growing vocabulary that is the driving mechanism behind the cognitive reorganization that leads to broad accuracy in the area of morphology. This is consistent with our post hoc correlations, which showed that vocabulary scores, but not age, predicted gains in treatment. It is worth noting that the PPVT-III does not uniquely assess verb vocabulary size; thus, as a predictive measure, it is not as specific as would be recommended by Li and Shirai. We also only administered the PPVT-III at the beginning of therapy, meaning that we cannot comment on the role that gains in vocabulary played in the treatment outcomes. We did measure the number of verbs produced on the probes at both pretest and posttest, and this did not appear to influence results. However, this is a very narrow assessment of vocabulary growth, given that the examiner provides multiple models of the target verb during probe administration. Future work should independently measure changes on verb vocabulary size to better understand the relationship between individual lexical items and the meaning of more abstract constructions (Ambridge et al., 2011, 2013).

Indeed, our results do suggest that a focus on meaning in grammatical learning is warranted (Goldberg, 1995) and that there may be some dissociation between input that supports initial accuracy and input that leads to acquisition of an abstract pattern. The finding that starting with hard verbs promotes generalization supports the hypothesis first advanced by Li and Shirai (2000) that children infer the importance of a morpheme by being exposed to cases in which the verb meaning and the morpheme meaning are in conflict. Saying that the children in the hard-first condition were better at inferring the meaning of past-tense *-ed* constructions than the children in the easy-first condition suggests an active, conscious process on the part of the hard-first children. This is not our intent. Rather, just as repeated pairings of similar-meaning words and morphemes (*drop + ed*, *wiggle + ing*) lead to acquisition of particular patterns via statistical learning mechanisms, repeated pairings of words

that occur less frequently with that particular morpheme may lead to different acquisition patterns. This is borne out in that Li and Shirai's hypothesis was developed as a result of a computational model, something that naturally lacks the ability to deduce, infer, or have insight independent of the distributional information made available to it.

The focus on meaning, as opposed to the grammatical pattern alone, is corroborated by Fey, Leonard, Bredin-Oja, and Deevy (2017). The greatest gains in their study occurred in the treatment condition that explicitly highlighted the time-based information associated with tense morphemes (i.e., *is/was jumping*) as opposed to a treatment condition that used the same activity to highlight lexical differences (i.e., *is jumping* vs. *is eating*). Further work on linking grammatical form and meaning for morphemes is warranted, and understanding of the meaning encoded by tense and agreement markers or other functional morphemes should not be assumed.

Our target words were not selected exclusively on the basis of meaning, but also on the basis of their phonological characteristics. One might imagine that the hard-first words also promoted gains in intelligibility or production of complex codas. We cannot rule that out. Complexity-first accounts of phonological acquisition argue that practice producing complex combinations of sounds and features will lead to gains in simpler combinations (Gierut & Champion, 2001). Most of this work has been done with word-initial *s-* clusters that violate common sonority-based principles with regard to syllable shape, but there is no reason to think that this would not be true for word-final inflections that also violate typical syllable shape rules. As with lexical semantics, evidence can be derived from modeling work. Computational models describing past-tense production emphasize that verbs that end in alveolar sounds may be treated as already marked for tense and thus be more likely to have a tense omission when the syllabic allomorph is required (Marchman et al., 1999). Exposure to cases in which past tense is marked overtly even when the lexical verb ends in *-t* or *-d* (e.g., *rest* → *rested*) may have a similar effect on the statistical regularities stored by the brain as regular exposure to cases in which the lexical semantics of the verb are not closely aligned with the semantics of the morpheme. These cases draw attention to the formation of the past-tense morphophonological pattern by illustrating the boundaries of the way the regular form is created.

This study was not designed to separate the contribution of phonological learning and semantic learning on past-tense acquisition, and further work will be required to determine whether both of these factors in combination or one of these factors alone is sufficient to lead to change. The work by Owen Van Horne and Green Fager (2015) indicated that both semantic and phonological factors contributed approximately equally to production accuracy, but because our results indicate that words that lead to improved production accuracy do not necessarily lead to improved acquisition, this may not be a relevant consideration. An item analysis of the particular words produced correctly and incorrectly at different points in treatment

and on different tasks may provide some insight into this question.

In the clinic, one might imagine that similar results should hold for thinking about acquisition of other morphemes (e.g., third-person singular paired with nonhabitual actions, progressive paired with stative and completive acts). It is worth noting that the complexity-first approaches to therapy are all building on principled contrasts that should lead to generalization via a particular mechanism. With this in mind, the recommendation is not that a clinician work outside of the zone of proximal development or on randomly selected unusual and rare targets but rather that the particular exemplars chosen for therapy should be selected to promote generalization. Although we might assume that the four children who never trained to criterion on any lists would be working outside of the zone of proximal development, the fact that half of these children made moderate gains at posttest calls this into question. It is not yet clear how to determine if a child is developmentally ready to learn regular past-tense *-ed* or, for that matter, most grammatical targets, although the presence of subject-verb combinations and noun plurals seem to be precursors for learning agreement (Pawłowska, Leonard, Camarata, Brown, & Camarata, 2008).

An alternative explanation of our findings is worth consideration. Plante et al. (2014) have emphasized the role of input variability on learning. That is, they have shown that hearing a morphological inflection with a wide variety of different verbs leads to better learning than hearing that inflection with a few carefully selected verbs. It is possible that the key mechanism at work is not highlighting meaning contrasts, but rather enhanced input variability. Given that the natural environment tends to emphasize verbs that are more well aligned with past tense, it may be that focusing on hard verbs first enhances the overall variability of the input rather than highlighting the meaning contrast. The converse argument could be made about Plante et al.'s approach: Emphasizing variability in therapy may necessarily ensure that the child is exposed to both well-aligned and poorly aligned verb + morpheme pairs, which could lead to better learning. Should it be shown that variability- and complexity-driven approaches rely on the same underlying principle, we would imagine that variability-driven approaches should be adopted, as they are easier to implement and likely less frustrating for the child and clinician because the focus is on variety rather than eliciting particular, less well-known verbs from the child.

To conclude, this study has shown that those factors that promote early accurate use may not be the best factors to emphasize to promote learning and generalization in the area of morphological acquisition. In support of a theory of grammar learning that promotes the role of meaning in acquisition, we have found that selecting exemplars to illustrate the boundaries of verb + morpheme pairings can influence treatment outcomes. Future work should further examine the mechanisms associated with this approach, including questions about the contribution of variability and phonology, and develop techniques to make this approach

more accessible to clinicians for a wider variety of morphemes and grammatical constructions.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant K23-DC 013291 awarded to Amanda Jean Owen Van Horne. Stimuli development and pilot work was made possible through a grant from the ASH Foundation, awarded to Amanda Jean Owen Van Horne. The paper benefited from discussion with Karla McGregor, Prahlad Gupta, and Gary Dell. Lauren Seemann and Diane Buffo played critical roles in the process of subject recruitment and retention and data management. Members of the grammar acquisition lab assisted with intervention provision, data collection, transcription, and analysis. School districts across Iowa supported the recruitment process, but some schools and university clinics went above and beyond. Thus, we also gratefully acknowledge Augustana College (Alli Haskill); St. Ambrose University (Elisa Huff); Grantwood and Mississippi Bend Area Education Agencies; and the Iowa City, Grimes, and West Branch Community school districts for their help with recruiting participants, identifying intervention providers, and providing space to test participants.

References

- Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, United Kingdom: Cambridge University Press.
- Ambridge, B., Pine, J. M., & Rowland, C. F. (2011). Children use verb semantics to retreat from overgeneralization errors: A novel verb grammaticality judgment study. *Cognitive Linguistics*, 22, 303–323.
- Ambridge, B., Pine, J. M., Rowland, C. F., Chang, F., & Bidgood, A. (2013). The retreat from overgeneralization in child language acquisition: Word learning, morphology, and verb argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4, 47–62.
- American Speech-Language-Hearing Association. (1997). *Guidelines for audiologic screening*. Retrieved from <http://www.asha.org/policy/GL1997-00199/>
- Anron, I., & Clark, E. V. (2011). Why *brush your teeth* is better than *teeth*—Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, 7, 107–129.
- Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 consortium. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development. Phase 2. Terminology. *PeerJ Preprints*, 4, e2484v1.
- Blom, E., & Paradis, J. (2013). Past tense production by English second language learners with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 56, 281–294.
- Bloom, L., Lifter, K., & Hafitz, J. (1980). Semantics of verbs and the development of verb inflection in child language. *Language*, 56, 386–412.
- Borovsky, A., & Elman, J. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language*, 33, 759–790.
- Camarata, S. M., Nelson, K. E., & Camarata, M. N. (1994). Comparison of conversational-recasting and imitative procedures for training grammatical structures in children with specific language impairment. *Journal of Speech and Hearing Research*, 37, 1414–1423.
- Cleave, P. L., Becker, S. D., Curran, M. K., Owen Van Horne, A. J., & Fey, M. E. (2015). The efficacy of recasts in language intervention: A systematic review and meta-analysis. *American Journal of Speech-Language Pathology*, 24, 237–255.
- Crystal, D. (1985). Some early problems with verbs. *Child Language Teaching and Therapy*, 1, 46–53.
- Dalal, R. H., & Loeb, D. F. (2005). Imitative production of regular past tense –ed by English-speaking children with specific language impairment. *International Journal of Language & Communication Disorders*, 40, 67–82.
- Dawson, J. I., Stout, C. E., & Eyer, J. A. (2003). *Structured Photographic Expressive Language Test—Third Edition*. DeKalb, IL: Janelle.
- Dodd, B., Crosbie, S., McIntosh, B., Holm, A., Harvey, C., Liddy, M., . . . Rigby, H. (2008). The impact of selecting different contrasts in phonological therapy. *International Journal of Speech-Language Pathology*, 10, 334–345.
- Dunn, L. M., & Dunn, D. M. (1997). *Peabody Picture Vocabulary Test—III*. Circle Pines, MN: AGS.
- Eisenberg, S. L. (2013). Grammar intervention: Content and procedures for facilitating children's language development. *Topics in Language Disorders*, 33, 165–178.
- Fey, M. E., Leonard, L. B., Bredin-Oja, S. L., & Deevy, P. (2017). A clinical evaluation of the competing sources of input hypothesis. *Journal of Speech, Language, and Hearing Research*, 60, 104–120.
- Gierut, J. A. (1989). Maximal opposition approach to phonological treatment. *Journal of Speech and Hearing Disorders*, 54, 9–19.
- Gierut, J. A., & Champion, A. H. (2001). Syllable onsets II: Three-element clusters in phonological treatment. *Journal of Speech, Language, and Hearing Research*, 44, 886–904.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.
- Hassink, J. M., & Leonard, L. B. (2010). Within-treatment factors as predictors of outcomes following conversational recasting. *American Journal of Speech-Language Pathology*, 19, 213–224.
- Johnson, B. W., & Fey, M. E. (2006). Interaction of lexical and grammatical aspect in toddlers' language. *Journal of Child Language*, 33, 419–435.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test—Second Edition*. Hoboken, NJ: Wiley.
- Leonard, L. B. (1975). Modeling as a clinical procedure in language training. *Language, Speech, and Hearing Services in Schools*, 6, 72–85.
- Leonard, L. B., Davis, J., & Deevy, P. (2007). Phonotactic probability and past tense use by children with specific language impairment and their typically developing peers. *Clinical Linguistics & Phonetics*, 21, 747–758.
- Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Berlin, Germany: de Gruyter.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 1–16.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.
- Marchman, V. A., Wulfeck, B., & Ellis Weismer, S. (1999). Morphological productivity in children with normal language and SLI: A study of the English past tense. *Journal of Speech, Language, and Hearing Research*, 42, 206–219.
- Marshall, C. R., & van der Lely, H. K. (2007). The impact of phonological complexity on past tense inflection in children with

- grammatical-SLI. *Advances in Speech-Language Pathology*, 9, 191–203.
- Mohammadzaheri, F., Koegel, L. K., Rezaee, M., & Rafiee, S. M.** (2014). A randomized clinical trial comparison between pivotal response treatment (PRT) and structured applied behavior analysis (ABA) intervention for children with autism. *Journal of Autism and Developmental Disorders*, 44, 2769–2777.
- National Commission for the Protection of Human Subjects of Biomedical & Behavioral Research & Ryan, K. J. P.** (1978). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research—the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research*. Washington, DC: U.S. Government Printing Office.
- Oetting, J. B., & Horohov, J. E.** (1997). Past-tense marking by children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 40, 62–74.
- Owen Van Horne, A. J., & Green Fager, M.** (2015). Quantifying the relative contributions of lexical and phonological factors to regular past tense accuracy. *International Journal of Speech-Language Pathology*, 17, 605–616.
- Pawlowska, M., Leonard, L. B., Camarata, S. M., Brown, B., & Camarata, M. N.** (2008). Factors accounting for the ability of children with SLI to learn agreement morphemes in intervention. *Journal of Child Language*, 35, 25–53.
- Perona, K., Plante, E., & Vance, R.** (2005). Diagnostic accuracy of the Structured Photographic Expressive Language Test: Third Edition (SPELT-3). *Language, Speech, and Hearing Services in Schools*, 36, 103–115.
- Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., . . . Burton, R.** (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology*, 23, 530–545.
- Posner, M. I., & Keele, S. W.** (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.
- Redmond, S. M., & Rice, M. L.** (2001). Detection of irregular verb violations by children with and without SLI. *Journal of Speech, Language, and Hearing Research*, 44, 655–669.
- Rimikis, S., & Buchwald, A.** (2015). Do morphophonological rules impact both regular and irregular verb inflection? Evidence from acquired morphological impairment. [Abstract]. *Frontiers in Psychology Conference Abstract: Academy of Aphasia 53rd Annual Meeting*. Retrieved from <http://doi.org/10.3389/conf.fpsyg.2015.65.00028>
- Rispens, J. E., & De Bree, E. H.** (2014). Past tense productivity in Dutch children with and without SLI: The role of morphophonology and frequency. *Journal of Child Language*, 41, 200–225.
- Rvachew, S., & Nowak, M.** (2001). The effect of target-selection strategy on phonological learning. *Journal of Speech, Language, and Hearing Research*, 44, 610–623.
- Rvachew, S., Rafaat, S., & Martin, M.** (1999). Stimulability, speech perception skills, and the treatment of phonological disorders. *American Journal of Speech-Language Pathology*, 8, 33–43.
- Schulz, K. F., Altman, D. G., & Moher, D.** (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, c332.
- Tomas, E., Demuth, K., Smith-Lock, K. M., & Petocz, P.** (2015). Phonological and morphophonological effects on grammatical development in children with specific language impairment. *International Journal of Language & Communication Disorders*, 50, 516–528.
- Vygotsky, L.** (1978). Interaction between learning and development. In M. Gauvain & M. Cole (Eds.), *Readings on the development of children* (2nd ed., pp. 34–40). New York, NY: Scientific.
- Weiler, B.** (2013). Verb selection and past-tense morphology: Crystal's criteria revisited. *Topics in Language Disorders*, 33, 152–164.
- Williams, K.** (1997). *Expressive Vocabulary Test*. Circle Pines, MN: AGS.
- Wulff, S., Ellis, N. C., Römer, U., Bardovi-Harlig, K., & Leblanc, C. J.** (2009). The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal*, 93, 354–369.

Appendix A (p. 1 of 2)

Target and Generalization Verbs Sorted Into Lists for Treatment

Table A1. Target verbs, along with predicted accuracy, are based on Owen Van Horne and Green Fager (2015).

Set	Verb	Predicted accuracy	Average predicted accuracy
Easiest (List 1)	trip	0.829	0.847
	cry	0.835	
	jump	0.841	
	scare	0.857	
	close	0.874	
Easier (List 2)	sneeze	0.778	0.799
	crawl	0.793	
	hug	0.794	
	climb	0.812	
	remember	0.817	
Easy (List 3)	hop	0.745	0.758
	bake	0.756	
	point	0.758	
	work	0.762	
	stretch	0.769	
Hard (List 4)	bark	0.708	0.722
	plant	0.716	
	paint	0.724	
	count	0.728	
	whistle	0.733	
Harder (List 5)	float	0.668	0.684
	scratch	0.679	
	snore	0.681	
	bounce	0.686	
	clap	0.706	
Hardest (List 6)	rest	0.549	0.613
	fish	0.609	
	rake	0.631	
	hum	0.633	
	listen	0.645	

Appendix A (p. 2 of 2)

Target and Generalization Verbs Sorted Into Lists for Treatment

Table A2. Generalization verbs sorted into lists for within-treatment testing along with predicted accuracy on the basis of Owen Van Horne and Green Fager (2015).

Set	Verb	Predicted accuracy	Average predicted accuracy
List 1	imagine	0.574	0.725
	dance	0.677	
	clean	0.730	
	stamp	0.787	
	play	0.857	
List 2	giggle	0.626	0.731
	whisper	0.681	
	smile	0.739	
	yell	0.794	
	kiss	0.815	
List 3	paddle	0.633	0.724
	yawn	0.686	
	stir	0.726	
	color	0.759	
	walk	0.817	
List 4	sail	0.642	0.736
	squish	0.692	
	roll	0.747	
	guess	0.763	
	slip	0.834	
List 5	exercise	0.545	0.716
	wiggle	0.706	
	wave	0.718	
	cough	0.777	
	carry	0.836	
List 6	growl	0.648	0.754
	believe	0.713	
	turn	0.758	
	discover	0.799	
	answer	0.853	

Note. All verbs were tested at pretest and posttest points.

Appendix B (p. 1 of 2)

Sample Treatment Script

A sample treatment script along with administration notes from the training manual, adapted here for length.

Sentence Imitation

Get the child's attention before saying the sentence. Administer each sentence, scoring only the underlined word. If the child gets the word correct on the first attempt, praise the child, being sure to model the target word in an inflected form (Great work, you said, "hopped." I heard the -d on the end of hopped, etc.). Try to make your praise specific, sincere, and meaningful to the child. Say the target word to equate the number of presentations across correct/incorrect responses. If the child does not inflect the underlined word with past tense, repeat the word in isolation, emphasizing the -t/-d at the end and ask the child to imitate you. Then attempt to have them say the target word in sentence again. On the second attempt, you may shorten the sentence to maximize success since it isn't scored and this is a teaching trial.

1. Across the finish line the boys hopped.
2. It was her birthday so we baked.
3. I couldn't find the book so she pointed.
4. We thought the toy was broken but then it worked.
5. The balloon filled with air and stretched.
6. My teacher pointed to the clock when I was late.
7. He worked so hard on the project.
8. The kangaroo hopped with her baby.
9. She baked cookies for all of her friends.
10. The chef stretched the pizza dough.

Prompting Hierarchy, illustrated with Sentence 1:

(a) Correctly inflected – sample response.

Great job! You said hopped.

I heard the -d on the end of hopped!

(b) Omitted – sample response

i. Say the word inflected in isolation/cue as needed to support use of -ed

Say hopped.

Let's put the -d on the end ... hop-d

ii. Say the word again in the sentence; you can shorten the sentence the second time to promote success

Now let's put it in the sentence. The boys hopped.

Note that there is typically one additional exposure for each sentence if the target is correct and two additional exposures per sentence if the target is incorrect.

STOP: Did the child produce -ed on the target verbs above 8/10 times on the first try? If yes, give Within Treatment C probe now. If not, continue.

Observational Modeling

Using any small characters that are of interest to the child, present each verb five times. Use five different characters. Present the same verb repetitively with the different subjects. Be sure the action is completed before describing the action (e.g., Here's Ernie. Make Ernie stretch. Look. Ernie stretched. Now here's Elmo. Make Elmo stretch. Look. Elmo stretched, etc.). The child is only required to listen, not to repeat or respond in anyway. Allow the child to make choices (e.g., which character goes first) to maintain interest.

Worked x 5
Pointed x 5
Hopped x 5
Baked x 5
Stretched x 5

Syntax Stories + Focused Stimulation

Read the stories below one by one. Follow each story with accompanying play with toys. Stories serve as a vehicle for extra models of the target verb in its inflected form and a means of supporting elicitation of these verbs via play. You may go off script during play. You may recast any verb from this list of verbs during play at any time. Tally the verbs as you recast them—at a minimum recast each verb three times. After you have recast a verb five times, actively avoid recasting that verb and move on to targeting other verbs on the list. If the child is not producing the target verbs, model the verbs heavily. Preferably

Appendix B (p. 2 of 2)

Sample Treatment Script

model the verbs inflected with –ed more than you model bare stem or other inflections. To maintain interest/participation on the part of the child you may do the stories in any order you would like.

Worked
Pointed
Hopped
Baked
Stretched

Friends of Different Sizes

Verbs: Point, Stretch

Toys: various groceries, a short character, a tall character, a way of having “shelves” (on/off the table, use of blocks)
Let’s meet Short Sheryl and Tall Timothy.

Short Sheryl is tiny. She crawls under things. She can’t reach things way up high. Tall Timothy is big. He steps over things. He gets tired of bending over all the time to pick up things down low.

Last week Short Sheryl and Tall Timothy went shopping. Tall Timothy looked at his grocery list.

“I need some rice way down there on the bottom shelf,” said Timothy. He **pointed** at the rice.

“I’ll get it for you,” said Sheryl. She picked up the rice and **stretched** up to hand it to Timothy.

Timothy **pointed** to a couple more things on the bottom shelf. He **pointed** to noodles, and he **pointed** to peas.

“I need some tomatoes,” said Sheryl. She **stretched** and **stretched** but couldn’t reach.

“I’ll get them for you,” said Timothy. He got down the tomatoes and gave them Sheryl.

“I need some milk, too,” said Sheryl. She **stretched** up to the second shelf! It was too high. She stood on tiptoes and **stretched** her arms as far as she could! Too far!

Timothy picked up the milk and gave it to Sheryl and then **pointed** to a yogurt container all the way at the bottom of the shelf. Sheryl handed it up to Timothy.

“We make a good team!” said Timothy.

“Yeah,” said Sheryl, “let’s work together!”

The Bake Shop

Verbs: Bake, Work

Toy food – rolls, bread, mixing bowls, pretend cakes and cookies.

Toy cleaning supplies and “shop” things.

A and B owned a bakery. A likes to do all the baking. B does other kinds of work. They **baked** things for customers.

Last week they **worked** on a different thing every day.

On Monday, A **baked** croissants. “I love rolling the dough,” said A.

B **worked** on the floors.

On Tuesday, B **worked** on cleaning display cases. “Everything looks shiny,” said B.

A **baked** bread. “It smells so good,” said A.

On Wednesday, A **baked** cookies. “I think cookies are yummy,” said A.

B **worked** at the cash register. “I sold lots of things,” said B.

On Thursday, A **baked** cakes. B decorated birthday cakes that A had baked! “It’s so fun decorating the tops,” said B.

On Friday, they both took a break. “We **worked** hard all week! It’s time to rest.”

What would you like to do in a bakery?

The Frog and the Bug

Verb: Hop

Froggy was hungry! He **hopped** around looking for something to eat! There was a bug next to him. He stuck his tongue out to get it, but the bug had already **hopped** away.

The frog **hopped** after it.

“I can get you,” thought the frog, “if I just hop fast enough.”

He **hopped** toward the bug and stuck his tongue out again.

The bug **hopped** up onto a stick! “Ha ha, I got away.”

The frog **hopped** even closer to the bug...

What do you think will happen next?

Tally

Worked
Pointed
Hopped
Baked
Stretched

Appendix C

Sample Past Tense Probe Item

The sample past tense probe item is adapted from Redmond and Rice, 2001.

Examiner: This is Sleepy Bear. He falls asleep all the time. He tries to pay attention to the show, but sometimes he gets so tired he just falls asleep. You can help Sleepy Bear. You watch very close, and if he forgets to watch, you tell him what happened.

Sleepy Bear: I want to pay attention, but sometimes I fall asleep. I'm glad you're here to help me.

[In items 1–5, Agents are Donkey and Pig]

1. **FISH**

Donkey wants to draw on his paper.

Donkey: Should I draw a dog or a cat? **DRAW a dog/a cat**

Pig wants to fish for something.

Pig: Watch me fish! **FISH**

I didn't see. Could you tell me about the show?

Examiner acts out again, while saying:

Donkey drew on his paper, and Pig _____ (fished).

Sample Scoring:

Correct: Child says fished or fished-id.

Wrong: Child says fishes or fish or is/was fishing.

Unscorable: Child says went fishing, caught a fish, caught a fish.

Child is unintelligible while producing target verb.

Child is off topic.

If response is unscorable, the item could be readministered one time to attempt to elicit a scorable response.