Research Article

Effects of a Complexity-Based Approach on Generalization of Past Tense *–ed* and Related Morphemes

Amanda J. Owen Van Horne,^a Maura Curran,^a Caroline Larson,^b and Marc E. Fey^c

Purpose: In a previous article, we reported that beginning treatment for regular past tense *–ed* with certain types of verbs led to greater generalization in children with developmental language disorder than beginning treatment with other types of verbs. This article provides updated data from that study, including the addition of data from 3 children, results from naturalistic language samples, and data from a third time point.

Method: Twenty 4- to 9-year-old children with developmental language disorder (10 per condition) were randomly assigned to receive language intervention in which the verbs used to teach regular past tense *–ed* were manipulated. Half received easy first intervention, beginning with highly frequent, telic, phonologically simple verbs, and half received hard first intervention, beginning with less frequent, atelic, and phonologically complex verbs. The design used a train-to-criterion approach, with children receiving up to 36 visits.

Performance was assessed using elicited production probes and language samples before intervention, immediately following intervention and 6–8 weeks later.

Results: Children in the hard first group showed greater gains on the use of regular past tense *–ed* in both structured probes (at immediate post only) and in language samples (at both immediate and delayed post). Gains attributable to therapy were not observed in untreated morphemes.

Conclusions: This study suggests that the choice of therapy materials, with an eye on the role that treatment stimuli play in generalization, is important for treatment efficacy. Clinicians should consider early selection of atelic, lower-frequency, phonologically complex verbs when teaching children to use regular past tense *–ed*. Further work expanding this to other morphemes and a larger population is needed to confirm this finding.

The statistical regularities of language are well documented (Monaghan, Shillcock, Christiansen, & Kirby, 2014). A deep appreciation of these regularities may prove critical for hypothesizing about language development and language use in typically developing children and in those with developmental disorders. The study described here was designed to further our understanding as to how the statistical patterns that link a lexical verb (e.g., *jump*) with an inflectional morpheme (e.g., *–ed*) might be manipulated to optimize the effects of intervention for children with developmental language disorder (DLD; also known as specific language impairment; Bishop, Snowling,

Thompson, Greenhalgh, & the CATALISE Consortium, 2016; Bishop, Snowling, Thompson, Greenhalgh, & the CATALISE-2 Consortium, 2017). Drawing inspiration from a computational model of tense acquisition (Li & Shirai, 2001), we showed that generalized use of past tense *–ed* was acquired more readily when we exposed children to verb stems that (a) appear relatively infrequently rather than frequently in past tense expressions, (b) end in complex phonological forms, and (c) carry a sense of incompleteness (e.g., *walk*, *show*, *flow*) compared to verbs used that have a strong sense of completeness or telicity (e.g., *drop*, *jump*, *close*; Owen Van Horne, Fey, & Curran, 2017). In that article, we only reported pre–post comparisons from elicited production data and provided data from only 17 children. In this update of the Owen Van Horne et al. study, we have added data from three children, examined past tense *–ed* use during language samples and from elicited production, and added data from a third delayed posttest point. These modifications enable us to examine more closely our conclusion that manipulation of statistical properties of language can inform and enhance decision making in grammatical intervention.

^aUniversity of Delaware, Newark

^bUniversity of Wisconsin–Madison

^cUniversity of Kansas Medical Center, Kansas City

Correspondence to Amanda J. Owen Van Horne: ajovh@udel.edu

Editor-in-Chief: Mary Alt

Received December 5, 2017

Revision received January 23, 2018

Accepted February 26, 2018

https://doi.org/10.1044/2018_LSHSS-STLT1-17-0142

Publisher Note: This article is part of the Special Issue: How Statistical Learning Relates to Speech-Language Pathology.

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

Factors Known to Influence Past Tense Accuracy

Phonology

It has long been acknowledged that morphemes take different surface shapes depending on the phonological form of the stem. For instance, regular past tense *-ed* can be realized as /d/, /t/, and /ɪd/, depending on the stem to which *-ed* is attached (consider *play* + *d*, *walk* + *t*, and *rest* + *ed*, respectively). Specifically, for regular verbs, the verbs that end in a vowel or a voiced consonant take a /d/ when inflected in the past tense (*play* + *d*, *carry* + *d*, *rub* + *d*, *bang* + *d*). Regular verb stems that end in a voiceless consonant take a /t/ for past tense inflection (e.g., *walk* + *t*, *jump* + *t*, *hop* + *t*). Finally, regular verb stems ending in /t/ or /d/ are inflected for past tense with the syllable /ɪd/ (e.g., *rest* + *ed*, *mend* + *ed*, *weed* + *ed*).

These different stem forms can lead to predictable differences in production accuracy (Berko, 1958). Children are most accurate when the inflection is phonologically simple (e.g., *played*). Next comes verbs that form homorganic consonant clusters in word-final position (e.g., *messed*, *buzzed*) when the inflection is added. Finally, typically developing children tend to be least accurate in producing the syllabic allomorph /ɪd/ (see work by Riches, 2015, suggesting opposite patterns in children with DLD). Although it seems rather obvious that the hierarchy of singletons and clusters is attributable to articulatory ease, changes in performance related to whether the clusters are phonotactically permitted or not suggest some influence of input frequency and some attention on the part of the child to the statistical regularities in how phonemes build words (Leonard, Davis, & Deevy, 2007). In addition, there is some speculation as to whether the syllabic allomorph is difficult because this construction is rare (Tomas, Demuth, Smith-Lock, & Petocz, 2015) or because the child assumes that words that end in [t] or [d] are already marked for past tense and need no further inflection (Marchman, 1997). Thus, we see that the phonological and frequency-based factors that affect how verbs form regular and irregular families are also relevant statistical properties of language.

Lexical Frequency

People tend to be more accurate at adding a grammatical marker to a word they know well (Stemberger & MacWhinney, 1986). Word frequency is a way of operationalizing how well children, in general, are likely to know words. When researchers have considered the role of frequency on accuracy, they have tended to look at two measures: how often do children hear the word in all of its forms (lemma frequency) and how often do children hear the word combined with the particular inflection of interest (inflected form frequency, e.g., past tense *-ed*; Blom & Paradis, 2013; Marchman, Wulfeck, & Weismer, 1999; Oetting & Horohov, 1997; Owen Van Horne & Green Fager, 2015). Drawing these numbers from databases of talk to children, rather than general databases, is important because words distribute differently in talk to adults compared to talk to children (e.g., Hills, 2013). The CHILDES database

contains one of the largest freely available data sets of child-directed speech, and tools are available to make frequency calculations straightforward (MacWhinney, 2000). Children are more accurate, in general, with words that are more frequent and even more accurate when the word occurs more often in the inflected form than in other forms (Owen Van Horne & Green Fager, 2015). The relative contributions and interaction between the frequency of the target inflection on a particular verb, on all verbs, and the appearance of other inflections on that particular verb are additional ways that statistical learning can influence language learning.

Aspect

Tense provides information about when events happen in time relative to the speaking event (present, past). For example, in the sentence, “I walked,” the walking must have occurred by the time of speaking for past tense to be appropriate. In contrast, aspect relates to how events extend over time. It includes information such as whether the action associated with the verb occurs instantaneously or with duration or information about how complete or incomplete the event expressed by the verb is (Comrie, 1976). In our study, we were concerned mostly with telicity. Telicity is the formal term for a continuum of completedness. Completed events with definite end points are said to be telic (e.g., *cross*, *jump*), and ongoing/incomplete events are said to be atelic (e.g., *walking* or *running*, as in a circle).

In actuality, events exist along a continuum ranging from telic to atelic, and the characterization of the verbs describing those events may interact with the rest of the sentence. For instance, the verb *walked* encodes an act that is low in telicity; it has no particular goal. It can be made more telic, however, by adding an adverbial goal, as in *she walked to the store*. Note that, in this example, if the actor fails to make it to the store, the act is not complete. That said, it is possible to classify and order verbs on the basis of telicity, such that some highly telic verbs such as *close* or *kill* are followed by verbs such as *jump* or *walk*. Stative verbs, such as *want* and *like* are generally considered the least telic verbs.

Children, like adult second-language users, are most accurate when the lexical aspect of the verb and the aspect associated with the inflectional morphology align (Blom & Paradis, 2013; Leonard, Deevy, et al., 2007; Shirai, 2015; Wulff, Ellis, Roemer, Bardovi-Harlig, & Leblanc, 2009). For instance, children are more accurate when inflecting telic verbs for past tense *-ed* (e.g., *closed*) than when inflecting atelic verbs (e.g., *walked*). As with the phonological information, the cause of this increase in accuracy is not clear. Semantic alignment may ease the cognitive load, increasing accuracy. That said, input frequency may also be one explanation. Telic verbs are more frequently heard with *-ed* (e.g., *missed* [the ball]) than with other inflectional forms such as third-person singular *-s* (e.g., *misses*) or *-ing* (e.g., *missing*), and verbs expressing ongoing actions more frequently occur with *-ing* (*walking*; Wulff et al., 2009).

Computational Models

In each of the three areas just described, the relative contributions of phonotactic probability, the pattern of the stem + the inflection, lemma and inflected verb frequency, and telicity can be difficult to determine through behavioral work. This is because many of the elements are confounded. For instance, the frequency with which a particular phonological pattern is heard is influenced by the frequency of the word, which is, in turn, affected by the semantic properties of the event. Two types of computational models, connectionist networks (Rumelhart, McClelland, & the PDP Research Group, 1986) and self-organizing maps (Li & Shirai, 2001), have contributed to our understanding of how the statistics of the input influence the learning process.

Most connectionist models for past tense acquisition build on Rumelhart et al.'s (1986) work, which demonstrated that statistical learning of phonological patterns could lead to learning the past tense. This was in contrast to the generally accepted view that past tense was formed either via direct retrieval of memorized words with the inflection already attached or through compositional rules, such as add *-ed* (Pinker, 2000). Plunkett and Marchman (1991) applied Rumelhart et al.'s model to children. They showed that the number of verbs contributing to each phonological family of verbs (e.g., *sang*, *rang*, *sank*, *drank*, *shrank*, vs. *went*) and the frequency of the individual verbs within that family interact to help predict which phonological patterns will be produced accurately and productively. These interactions are difficult to predict based on a casual review of the phonological families because of the interactions between type and token frequencies over developmental time. Importantly, the models suggested that children initially overregularize and apply the rule of adding *-ed* broadly as the variety of verbs that they hear expands rapidly.

Self-organizing maps (Kohonen, 2001) are another type of computational model that provides a visualization of the relationships between highly complex information. Using a self-organizing map, Li and Shirai (2001) demonstrated that lexical aspect interacts with the frequency with which a word is heard in a particular inflected form. They trained their model with semantic and phonological information presented in a developmentally ordered fashion. The model grouped together words that ended the same, which also led to groupings based on similar aspectual properties.

Children's early accurate productions are with highly aligned verb + morpheme combinations (e.g., *bang* + *ed*, *run* + *s*), but they eventually are able to apply a regular morpheme to any new verb. Indeed, we have already reported that children are more accurate at inflecting telic verbs with past tense *-ed* than atelic verbs. The model suggested that children become productive at the point at which they are exposed routinely to cases in which the tense marker is presented with a verb from a different aspectual pattern than expected. So generalization starts to occur when children hear atelic verbs with *-ed* or stative verbs

with *-ing*, even if this is not the most common form of input provided.

In both of these cases, the computational model provided insights into the learning process that might be difficult to discover through reflection or a review of existing data. Despite different ways of representing language forms, these models both showed that the increasing diversity of verbs that children are exposed to as they develop drives critical transformations that allow for generalization of the regular inflectional rule to new cases. This is influenced both by the diversity of phonological patterns present in the connectionist models and the diversity of semantic meanings present in the self-organizing maps.

Role of These Factors in Target Selection

When attempting to design treatment for children with difficulty in using tense and agreement markers, the literature provides very little guidance on the appropriate target words to select. One might infer from the studies that examine which factors promote more accurate productions that we should reduce the phonological complexity, increase lexical word frequency, and align the tense morpheme with lexical aspect in order to best elicit correct productions. This is the general consensus of expert opinion in the area (Crystal, 1985; Weiler, 2013), and it is consistent with approaches that suggest that lots of "right practice" leads to better outcomes (Rvachew, 2005; Warmington, Hitch, & Gathercole, 2013). To draw from another domain, this would be like exposing individuals to bluebirds, cardinals, robins, and wrens in order to teach them the category "bird" (Borovsky & Elman, 2006; Posner & Keele, 1968). In fact, one recommended approach in treatment is to pick a subset of "easy" verbs to focus on that are high frequency, telic, and relatively easy phonologically to inflect and practice them until the child is highly accurate. As the child masters those easy verbs (Weiler, 2013), the difficulty of the verbs chosen for therapy practice is gradually increased. This is consistent with target selection approaches in which the goal is early success and stepwise accuracy within the zone of proximal development (e.g., Rvachew & Nowak, 2001).

Work with computational models of morphophonological learning suggests that an alternative strategy warrants careful consideration. Using this strategy, verbs targeted in therapy would not be the easiest, the most frequently used, or the best semantically aligned with the target morpheme (e.g., past tense + *ed* on verbs high in telicity). In contrast, if we are seeking consistent generalized use of morphosyntactic forms, we might adopt a strategy that would involve exposure to words that are less common, take allomorphs that are less prototypical (e.g., /ɪd/ rather than /t/ or /d/), and have telicity ratings that align poorly with the target morpheme (e.g., atelic verbs to learn the past tense.) Much like knowing why a penguin is a bird, even though it does not fly, requires fundamental understanding of the category *bird*; in the same way, inflecting a verb that expresses an ongoing event, like *rest* or *wiggle*, with *-ed*

demonstrates broader mastery of the past tense. Similar complexity-based approaches that select targets on the basis of promoting widespread generalization have been adopted in the areas of phonology (least phonological knowledge/complexity-based approaches; Gierut, 2001), aphasia (complexity account of treatment efficacy; Thompson & Shapiro, 2007), and anomia (semantic complexity; Kiran & Thompson, 2003). These complexity-based approaches have also been argued for abstractly in language therapy using general learning principles (Alt, Meyers, & Ancharski, 2012). It is important to point out that changing the exposure in therapy does not change the way English works in the rest of the child's life. Thus, the model's assumption that children are initially learning via exposure to the "easy" forms is not violated. What could be happening in therapy, though, is increased and systematic exposure to the forms that are not common in everyday life, with the goal of accelerating generalization.

Questions

Given these two contrasting approaches, we designed a study to ask the following questions:

Q1. Does initial exposure in treatment to verbs that are phonologically simple, highly frequent, and telic (hereafter, "easy" verbs) or verbs that are phonologically complex, less frequent, and atelic (hereafter, "hard" verbs) lead to greater accuracy for regular past tense *-ed* morphemes on structured probes and on language samples?

Q2. To what extent are any gains documented at the conclusion of treatment maintained over time?

Q3. Does initial exposure in treatment to easy verbs or hard verbs lead to better accuracy for untrained morphemes in language samples?

There are four major differences between analyses performed in our original study and this update. First, three participants who did not take part in the original study (Owen Van Horne et al., 2017) were included in the current investigation. Second, in this study, we added analyses on data at a point approximately 6–8 weeks after conclusion of treatment. Third, data for the current report included speech from spontaneous language samples as well as the original elicited production probes. Finally, in the updated study, we examined the use of morphemes other than past tense *-ed*.

Method

Participants

Twenty children with DLD participated in this study. Seventeen of these children constituted the participants of the original study (Owen Van Horne et al., 2017). Three additional children became available and were enrolled in the study after the report of the original study was submitted for publication. All participants (ages 4;0–10;0 [years; months]) met criteria for DLD. They received a standard

score below 94 on the Structured Photographic Expressive Language Test–Third Edition (SPELT-3; Dawson, Stout, & Eyer, 2003; see Perona, Plante, & Vance, 2005, for standard score cutoffs) and a standard score above 85 on the matrices subtest of the Kaufman Brief Intelligence Test–Second Edition (Kaufman & Kaufman, 2004). The use of a cutoff of 94 on the SPELT-3 can be questioned, because 10 of our participants were age 6;0 or older and thus were older than the participants of the Perona et al. study on which our criterion was based. Of the 10 children ages 6–10 who were included, eight had scores below 85. The two older children with SPELT-3 scores above 85 were receiving speech-language services through the public schools at the time of our study. Children passed a hearing screening and had no history of autism or other psychiatric or neurological disorders per parent report. All children enrolled in the study also demonstrated less than 40% use of regular past tense *-ed* on structured probes modeled after Redmond and Rice (2001), the ability to produce the target verbs on the probes at least 50% of the time at pretest, and the ability to use word final *-t* and *-d* in monomorphemic contexts with more than 80% accuracy. Children also had to be producing subject–verb combinations reliably (Hassink & Leonard, 2010) and have sufficient intelligibility to participate in testing procedures. The Peabody Picture Vocabulary Test–III (Dunn & Dunn, 1997) and the Goldman-Fristoe Test of Articulation–Second Edition (Goldman, Fristoe, & Williams, 2000) were administered for additional descriptive information (see Table 1).

Participants were assigned at random to either the hard first intervention, which began treatment with verbs anticipated to make past tense usage difficult, or the easy first intervention, which began treatment with verbs anticipated to make past tense usage easy. There were no significant differences between groups on the demographic information reported in Table 1, except for the use of

Table 1. Demographic characteristics of participants.

Demographic variable	Easy first	Hard first	<i>t(df)</i>	<i>p</i>
<i>n</i> (female)	10 (3)	10 (3)		
Age in months	64.7 (23.76)	70.8 (22.77)	<i>t</i> (18) = 1.08	.29
SPELT-3 SS	79.33 (10.16)	75.6 (13.28)	<i>t</i> (17) = 0.56	.58
KBIT-2 SS	102.3 (12.17)	100.4 (8.08)	<i>t</i> (18) = 0.82	.42
PPVT-III SS	101.56 (9.23)	97.6 (10.14)	<i>t</i> (18) = 1.46	.16
EVT SS	96.44 (9.85)	91.0 (13.34)	<i>t</i> (18) = 1.24	.23
GFTA-2	92.11(16.54)	94.20(16.09)	<i>t</i> (16) = 1.01	.33
Percent accuracy <i>-t/-d</i>	97.78 (4.41)	96.0 (5.48)	<i>t</i> (18) = 2.12	.048

Note. One child in the easy first condition was 10;0 years old when enrolled and therefore technically outside of the norms of the SPELT-3. Using 9;6- to 9;11-year-old norms, the child scored below the first percentile. Her score is not included in the average above. SS = standard score; SPELT-3 = Structured Photographic Expressive Language Test–Third Edition; KBIT-2 = Kaufman Basic Intelligence Test–Second Edition, Matrices subtest; PPVT-III = Peabody Picture Vocabulary Test–III; EVT = Expressive Vocabulary Test; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition.

word final *-t* and *-d*. For this measure, the easy first group was slightly more accurate at the use of final *-t* and *-d* than the hard first group, though there were no differences in overall articulation abilities as measured by the Goldman-Fristoe Test of Articulation–Second Edition. For a more complete description of recruitment procedures, please see Owen Van Horne et al. (2017).

Procedure

Treatment

All children received treatment for regular past tense *-ed*. During treatment, the particular verbs used to deliver treatment varied systematically across conditions. Half of the children, selected at random, participated in “hard first” intervention in which they began intervention with words ranked as unlikely to be inflected accurately according to Owen Van Horne and Green Fager (2015). The other half of the children, enrolled in the “easy first” condition, began intervention with words ranked as more likely to be inflected accurately.

Verb Selection

Verbs were selected from a set of 60 verbs analyzed in Owen Van Horne and Green Fager (2015) for ease of inflection. As described in the introduction, word-level factors, such as the phonological composition of the stem, the frequency of the inflected word from CHILDES (MacWhinney, 2000), and the completed nature of the event described by the verb as rated by adult English speakers, were used to predict how likely a child was to accurately inflect that particular verb. These 60 verbs were initially divided into two sets of 30 verbs each, such that the two sets were approximately equal in the likelihood of being inflected accurately. One set of verbs was designated the treatment set, and the other was designated the generalization set. The treatment set was further divided into six lists of five verbs each. Each set was structured in terms of difficulty so that the first set contained the five easiest verbs and the last set contained the five hardest verbs. In the generalization set, verbs were subdivided into six sets of five verbs that were approximately equal in difficulty, such that no set was significantly different from another set. Table 2 provides further description and examples of properties influencing verb complexity.

Intervention Procedure and Fidelity

Intervention was designed to reflect a variety of intervention practices commonly used with this age range, such that a mix of drill, modeling, and naturalistic approaches were employed (Cleave, Becker, Curran, Owen Van Horne, & Fey, 2015; Eisenberg, 2013; Owen Van Horne et al., 2017). The goal was to keep the focus on differences in target selection and to serve a range of children of different ages and with different learning styles, rather than to emphasize the choice of a particular approach to providing intervention.

A brief sketch overview of methods is provided here, with more details available in Owen Van Horne et al. (2017).

Within each session, one set of five verbs was targeted in drill practice, based on work by Finestack and colleagues (Finestack, 2014; Finestack & Fey, 2009). Each session included 10 sentence imitation items (two per verb, one sentence medial and one sentence final) presented in drill fashion with corrective feedback (e.g., “That’s right. You put ‘d’ on the end of jump—jumped” or “Next time let’s put the *-d* on the end, jumped”). In addition, the clinician provided 25 observational modeling examples (five per verb; Leonard, 1975), at least 15 exposures (three per verb) within a syntax story (Leonard, Camarata, Brown, & Camarata, 2004), and between 15 and 25 regular past tense *-ed* recasts during play-based focused stimulation (three to five per verb; Fey, Cleave, Long, & Hughes, 1993; Leonard et al., 2004).

One question in the original study was the extent to which selecting hard first verbs influenced rate of progress in therapy, not just final accuracy levels. This hypothesis was not borne out, but it influenced the study design. We had hypothesized that children beginning in the hard first condition might demonstrate slow initial performance and then speed up as they became better at generalizing. Thus, we used a train-to-criterion approach. Children received treatment on a single verb set (e.g., easy first: Verb List 1; hard first: Verb List 6) until they produced the sentence imitation and elicited production probes with at least 80% accuracy. At this point, they moved on to the next verb list (e.g., easy first: Verb List 2; hard first: Verb List 5). Time in therapy varied depending on how fast the child trained to criterion on each list, ranging from 13 visits for one child who completed all six lists rapidly to the maximum allowed of 36 visits for children who never completed all lists. After the 36th visit, children were dismissed, and posttesting was completed regardless of which verb list the child was on. Children in both conditions were exposed to the same verbs, albeit in different orders, if they completed all of the verb lists. Children completed between zero and six verb lists in the time provided, with no differences across condition in the number of verb lists completed. Scheduling of visits was done at the convenience of the family and ranged from one to three visits per week, with occasional longer breaks for holidays or illness.

Intervention was provided across the state of Iowa by research assistants with credentials ranging from undergraduate students in communication disorders and early education to licensed speech pathologists with the certificate of clinical competence. We maintained intervention fidelity by providing direct training, written scripts for each session, and phone and in-person coaching based on intervention audio recordings or direct observations. Between 25% and 100% of sessions were checked for fidelity for any particular child. Given that additional children are reported here, we updated the fidelity measures as well. Examiners were highly reliable at providing the sentence imitation (session accuracy: easy first, $M = 100\%$, $SD = 0\%$; hard first, $M = 97.7\%$, $SD = 4.9\%$), observational modeling (session accuracy: easy first, $M = 92.5\%$, $SD = 9.8\%$; hard first, $M = 93.4\%$,

Table 2. Properties that influence verb complexity (Owen Van Horne & Green Fager, 2015).

Property	Definition	Examples
Frequency (inflected form)	Based on CHILDES (MacWhinney, 2000), the number of times on average children hear a given verb + <i>ed</i> in their linguistic environment. Highly frequent in the + <i>ed</i> form increases accuracy.	Highly frequent: <i>closed, scared, walked</i> Less frequent: <i>exercised, rested, imagined</i>
Frequency (lemma)	Based on CHILDES (MacWhinney, 2000), the number of times on average children hear a given verb in any form in their linguistic environment. Highly frequent in any form reduces accuracy.	Highly frequent: <i>close, remember, turn</i> Less frequent: <i>snore, paint, squish</i>
Telicity	An adult rating of 1–4 on whether a verb has a clear end point or is an ongoing activity. Clear end points are more accurately inflected.	Highly telic: <i>close, answer, sneeze</i> Less telic or atelic: <i>rest, work, listen</i>
Phonological form	2 primary sources of variance in <i>-ed</i> production accuracy: word final obstruents and word final alveolars both make accurate inflection less likely	Ends in an obstruent: <i>close, slip, imagine</i> Ends in an alveolar: <i>close, rest, float</i> Ends in a continuant: <i>play, whistle, answer</i>

$SD = 9.7\%$), and syntax stories (session accuracy: easy first, $M = 89.1\%$, $SD = 12.6\%$; hard first, $M = 89.2\%$, $SD = 13.1\%$) as scripted and did not differ across conditions ($ps > .16$). Recast rates were poorer (percentage of sessions in which all five target verbs were recast at least three times: easy first, $M = 74.1\%$, $SD = 22.6\%$; hard first, $M = 72.1\%$, $SD = 31.4\%$; percentage of sessions in which at least 15 recasts were provided: easy first, $M = 94.8\%$, $SD = 10.3\%$; hard first: $M = 85.3\%$, $SD = 25.6\%$) but did not differ across conditions ($ps > .28$). Cases where recast targets were not accomplished were largely attributable to failure to elicit a platform utterance that included the target verbs, rather than missed recast opportunities on the part of the clinician. In all cases, better fidelity favored the easy first condition.

Outcome Measures

Children's performance on regular past tense *-ed* was assessed using structured probes and narrative retells at three time points: immediately prior to intervention beginning, immediately after intervention concluded, and 6–8 weeks after intervention was concluded. Because children's progress through therapy varied and the treatment visit frequency varied from one to three times per week, there is considerable variability between testing time points. For the easy first group, there was an average of 153.7 days ($SD = 29.26$ days) between pre- and posttesting and an average of 41.9 days ($SD = 17.9$ days) between posttesting and delayed posttesting. For the hard first group, there was an average of 117.7 days ($SD = 45.37$ days) between pre-post testing and an average of 51.1 days ($SD = 15.5$ days) between posttesting and delayed posttesting. The easy first group had a longer lag between pre- and posttesting, $t(19) = 2.10$, $p = .049$, whereas the hard first group tended toward a longer lag between posttesting and delayed posttesting, $t(19) = 1.66$, $p = .11$. These differences do not reflect the number of visits and have more to do with vagaries in

scheduling due to holidays, illnesses, and sports schedules than progress in treatment.

Past Tense Probes

Past tense probes consisted of 60 elicited production items modeled after Redmond and Rice (2001). All verbs in both the treatment and generalization sets were tested at each of the three time points. Children watched a short puppet show depicting two characters doing two events. The experimenter described one event in the simple past using an irregular past verb, and the child was asked to complete the sentence by describing the second event using the target verb in the regular past tense form ("Elmo wore a hat and Ernie..."). Experimenters were allowed to reprompt one time for off-topic responses or responses using a nontarget verb. Only target verbs were scored. Responses were scored as correct, omitted, or unscorable. Uses of alternative inflections (*was -ing* or third-person singular *-s*) were considered unscorable.

Frog Stories

In general, we elicited three frog story retells at each time point, though occasionally, there are missing data. Table 3 shows the average number of stories collected for each group at each time point. Scripts available from the Systematic Analysis of Language Transcripts website were used to elicit the stories *Frog Where Are You?*, *One Frog Too Many*, and *A Boy, a Dog, a Frog, and a Friend* (Mayer, 1970, 1975; Mayer & Mayer, 1969; Miller & Iglesias, 2012). The child listened to the story while looking at the corresponding storybook pages on a laptop computer screen. The examiner clearly indicated to the child that she was not attending to the story (e.g., not in the room, working on another task) and then asked the child to retell the story while using the hardcopy wordless picture book as a reference. If the child had difficulty in starting or completing the task, the examiner prompted

Table 3. Language sample information.

Values calculated	Easy first			Hard first		
	Pre	Immediate post	Delayed post	Pre	Immediate post	Delayed post
Mean <i>n</i> of stories	2.9 (0.31)	3 (0)	3 (0)	3 (0)	2.6 (0.96)	2.7 (0.67)
Mean <i>n</i> of utterances	86.1 (23.87)	86.9 (20.57)	94.3 (28.28)	94.6 (32.35)	82.1 (32.99)	91.9 (34.69)
Mean MLU	5.65 (1.07)	5.92 (1.12)	6.14 (0.73)	5.94 (1.40)	6.02 (1.81)	5.99 (1.74)
Past <i>-ed</i> correct/opportunities	81/206	120/256	121/257	91/248	143/203	220/267
Present correct/opportunities	118/126	104/113	114/119	140/149	101/117	115/119
Other past forms correct/opportunities	246/398	323/438	384/492	315/449	294/374	320/397

Note. MLU = mean length of utterance.

the child to continue (e.g., “Can you tell me something about every page?”).

Retells were transcribed according to the Systematic Analysis of Language Transcripts protocols (Miller & Iglesias, 2012) from audio recordings. Utterances were segmented into terminal units (T-units), which include a main clause and its subordinate clauses, and coded for bound morphology use. Given that the model stories were in past tense, we assumed that missing morphemes were past tense unless there was evidence from the child’s surrounding utterances to suggest otherwise. From the transcripts, we extracted (a) the number of complete and intelligible utterances, (b) the mean length of utterance (MLU) in words, and (c) the number of times that regular past tense *-ed*, other past tense markers (*was*, *were*, irregular past), and other tense-related markers (*am*, *is*, *are*, third-person singular *-s*) were used and omitted.

Reliability

Twenty percent of probes and 20% of the frog stories (two children per condition at all three time points) were randomly selected and independently retold and rescored by research assistants blind to condition and the hypotheses of the study. For the probe data, point-by-point reliability for past tense *-ed* use on the target words averaged 90% agreement at pretest, 89% agreement at immediate posttest, and 85% agreement at delayed posttest. For the frog story retells, intraclass correlations were computed for the following measures reported in this article: MLU in words, number of utterances, and percent correct for past tense *-ed*, present tense composite (third-person singular, *am*, *is*, *are*), and other past composite (irregulars, *was*, and *were*). Intraclass correlations ranged from .93 (MLU) to .98 (other past composite), with the exception of present tense, which had an intraclass correlation of .76. This is largely attributable to the fact that present tense forms were rare or completely nonexistent in many stories. Thus, it was common to have small differences in scoring that led to large differences in values, such as zero or one present tense morpheme opportunities identified. Similarly, there were instances in which two out of three correct present tense forms were recorded by one transcriber and three out of four were judged correct by a second transcriber.

Planned Analysis

In principle, random assignment should yield roughly equivalent groups on a wide variety of measures. However, in practice, with small group sizes, one cannot assume this. Thus, we initially tested for group equivalence on the diagnostic and descriptive measures that we collected using *t* tests. We also assessed group equivalence for language sample measures (number of utterances and utterance length) using a linear mixed model with group and time as fixed effects and participant and story as random effects.

For the questions of interest, mixed model logistic regressions were used as the analysis method, with group and time point as the fixed effects (Baayen, 2008). In the group effect, easy first was the reference variable. Because time point was a three-level (pre/post/delayed post) categorical effect, we initially ran the model with pretest as the reference variable, allowing us to compare pretest to posttest and pretest to delayed posttest. To directly compare posttest to delayed posttest, it was necessary to rerun the analysis using posttest as the reference variable. Using a logistic regression has the advantage of modeling the underlying distribution of the data, in which use of a particular morpheme is either correct or incorrect. Mixed effects allow us to mathematically adjust for the variability associated with the individual child and the elicitation context (verb, story). In addition, the model corrects for the fact that some children and some time points will have more opportunities for particular morphemes than others and gives more weight in the outcome to those cases with more data. In this way, we avoid the need to omit subjects or combine elicitation contexts. Random effects differed across the two data sets. For the language samples, we added participant and story as random intercepts; for the probes, we added participant and verb. Because participant is a random effect, the model accounts for individuals who “start strong” with many opportunities or high rates of accuracy initially. Likewise, if a particular verb or story are generally easier or harder than the others, this is taken into account mathematically. In both cases, we initially considered accuracy for the treated morpheme, past tense *-ed*. In the language samples, we also considered accuracy for all other past markers (irregular past, *was*, *were*) and for other tense

markers (contracted and uncontracted *am*, *is*, and *are*, and third-person singular *-s*).

Results

Structured Probes

We initially considered whether assignment to easy first or hard first treatment conditions influenced the likelihood that children would accurately produce past tense on structured probes. Results revealed a main effect for time indicating gains over pretest at both immediate posttest ($Z = 7.59, p < .0001$) and delayed posttest ($Z = 6.59, p < .0001$). Although there was no main effect of treatment condition ($Z = 0.55, p = .58$), there was an interaction between group and time indicating that, as previously reported in Owen Van Horne et al. (2017), the hard first group was more likely to use an accurate past tense inflection than the easy first group at both immediate posttest ($Z = 2.87, p = .004$) and delayed posttest ($Z = 3.50, p = .0005$). That is, children in the hard first group showed average gains in proportion correct of .35, and children in the easy first group showed average gains of .18, as shown in Table 4 and Figure 1. Shifting the reference value for time point indicated that there were no significant differences between the two posttest time points, either as main effects, $p > .25$, or as an interaction, $p > .48$. This suggests that nearly all of the observed gains had been obtained by the conclusion of treatment. Continued growth after the conclusion of therapy on the structured probes appears evident visually but is not reliable statistically. See Table 5 for the regression table.

Frog Stories

We first considered whether changes in the sample as a whole might have occurred, with particular attention to group differences that might influence our interpretation of the morphological gains. Sample size (number of utterances) did not differ for group or time point, nor was there an interaction between group and time point (all $ps > .24$). Mean length of utterance, measured in words, also did not differ across groups ($Z = 0.50, p = .61$). Samples collected at the delayed posttest were significantly longer than samples collected at pretest ($Z = 2.58, p = .01$), with immediate posttest samples falling in between ($Z = 1.48, p = .14$). There was no interaction between group and time point (Time2: $p = .72$, Time3: $p = .22$), indicating that both groups had similar gains in MLU in words over the 6 months that measurements were taken, something that would be consistent

with both general development and participation in focused stimulation.

Next, we considered whether the likelihood of producing past tense *-ed*, the targeted morpheme, was predicted by group and time point. Accuracy at each time point is shown in Table 4. Although there were no main effects for group ($Z = 0.192, p = .85$) or time point (immediate post: $Z = 1.51, p = .13$; delayed post: $Z = 1.50, p = .13$), there was an interaction between group and time (immediate post: $Z = 3.77, p < .0002$; delayed post: $Z = 5.98, p < .0001$). This interaction indicates that children in the hard first condition were more likely to produce regular past tense *-ed* correctly at the immediate and delayed post measurement points than were the children enrolled in the easy first condition. Shifting the reference value for time point allows direct examination of possible gains between immediate and delayed post time points also and revealed that the hard first group, but not the easy first group, demonstrated continued growth in past tense accuracy in language samples ($Z = 2.19, p = .03$) after the conclusion of treatment, as illustrated in Figure 2 and Table 4. The regression model is shown in Table 6.

We also examined the possibility that treatment for regular past tense *-ed* might have improved the use of other related past tense morphemes (irregular past, *was*, *were*) and present tense morphemes (*am*, *is*, *are*, third-person singular *-s*). Table 4 and Figure 2 also illustrate the findings from other past tense and present tense morphemes. For other past morphemes, no main effects or interactions associated with treatment condition emerged ($p > .30$). There was a main effect for time, with significant gains over pretest evident at both immediate posttest ($Z = 3.34, p = .0008$) and delayed posttest ($Z = 4.98, p < .0001$). Shifting the reference value indicated that performance at immediate and delayed posttest was not significantly different across time points ($Z = 1.6, p = .10$). Similarly, we examined whether treatment for regular past tense led to improvement in the use of present tense forms. There were no main effects or interactions (all $ps > .29$), indicating no significant changes in the use of present tense attributable either to the passage of time or to treatment condition.

Discussion

This article reports follow-up data and analyses for an early efficacy study (Owen Van Horne et al., 2017). With the exception of three participants added for the current

Table 4. Proportion accurate use of morphemes on each task at each time point.

Task	Morpheme	Easy first			Hard first		
		Pretest	Posttest	Delayed post	Pretest	Posttest	Delayed post
Probes	Regular past	.21 (.14)	.39 (.27)	.36 (.23)	.23 (.13)	.57 (.25)	.56 (.27)
Language sample	Regular past	.39 (.25)	.45 (.26)	.46 (.26)	.40 (.25)	.68 (.19)	.77 (.18)
	Other past	.56 (.26)	.70 (.18)	.74 (.22)	.69 (.17)	.73 (.22)	.78 (.22)
	Present	.87 (.17)	.80 (.32)	.91 (.17)	.92 (.10)	.88 (.13)	.99 (.02)

Figure 1. Proportion correct use of *-ed* on elicited production probes.

investigation, the participants were the same as the original study. It is not surprising, then, that comparisons between groups immediately following treatment confirmed the previously reported finding that the hard first group had higher accuracy rates than the easy first group. Data from the language samples at immediate and delayed posttest and probe data from the delayed posttest were not included in the original study. These results also illustrate that children who received the hard first condition were more accurate at the use of past tense than children who received the easy first condition. Differences in the groups' average outcomes ranged from a difference in average probe gains for proportion correct of .17 to a difference of the average gain for proportion correct on language samples of .31, both favoring the hard first group. On the structured probes, most of the gains appear to have been obtained by immediate posttest and then maintained 4–8 weeks until delayed posttest. On the language samples, higher

accuracy levels for the hard first group were observed at immediate posttest (hard first: $M = .68$, easy first: $M = .45$), and then even higher accuracy levels were observed at the delayed posttest (hard first: $M = .77$, easy first: $M = .46$), suggesting ongoing generalization over time. These ongoing gains were restricted to the trained morpheme past tense *-ed* and did not appear to generalize to other grammatical markers such as third-person singular *-s*, irregular past, or forms of *be*, such as *am*, *is*, *are*, *was*, or *were*. Groups were well matched at pretest, with the exception of the use of final consonants, which favored the easy first group. Thus, this study confirms that concentrated exposure to verbs that are “hard,” that is, verbs that are atelic, have complex phonological forms, and are less frequently inflected in the past tense, can lead to better posttest performance than exposure to verbs that are “easy.” This is consistent with the findings of other research programs that have taken a complexity-based approach to learning (Gierut, 2001; Thompson & Shapiro, 2007). The developmental and most intuitive model of goal selection (i.e., teach the easiest forms first) may not lead to optimal generalization of newly acquired morphosyntactic forms.

Less intuitive plans that are sensitive to statistical properties of input and learning may yield better clinical outcomes. The predictions of the computational models require us to focus on the role of the statistical information in the input. Recall that the initial presentation of a few highly frequent verbs paired with well-aligned morphemes, followed by systematic presentation of less common verbs, leads to development in the model. Our role as therapists is to provide enhanced exposure to past tense *-ed* on verbs that are atelic, phonologically complex, and less frequent. It is helpful to remember that children live in the world, not the therapy room. The general linguistic environment likely provides sufficient exemplars of the highly frequent cases to serve as the foundational presentation of well-aligned verb and morpheme pairs that the models rely on initially. Treatment then diversifies children's exposure

Table 5. Regression model for structured probes examining the influence of condition assignment and time on past tense *-ed* production.

Random effects	Variance	SD		
Verb	0.41	0.64		
Subject	1.26	1.12		
Fixed effects	β	SE	Z	p
Intercept				
Ref values: pretest, easy first	-1.91	0.39	-4.92	< .0001***
Immediate post	1.23	0.16	7.59	< .0001***
Delayed post	1.07	0.16	6.59	< .0001***
Hard first	0.30	0.53	0.55	.58
Immediate post * HF	0.64	0.22	2.87	.004*
Delayed post * HF	0.78	0.22	3.50	.0005**

Note. $N = 3,274$; 60 verbs, 20 subjects. HF = Hard First; AIC = Akaike Information Criterion, 3,500.1; BIC = Bayesian Information Criterion, 3,548.9; $df = 3,266$; LL = Log Likelihood, 1,742.1.

* $p < .01$. ** $p < .001$. *** $p < .0001$.

Figure 2. Proportion correct use of finite morphemes in frog story retells. Pre = pretest; Imm Post = Immediate posttest; Del Post = delayed posttest.

to the less common examples, so that children can store sufficiently strong representations of these words to trigger generalization of the morpheme to a wide range of verbs and contexts.

These findings highlight the role that computational models combined with treatment studies can have in better understanding language learning mechanisms for all children. Although our labs have described the interactions between phonological and semantic factors with regard to verb use (Johnson & Fey, 2006; Owen Van Horne &

Green Fager, 2015), the computational model by Li and Shirai (2001) first suggested to us that selecting hard first type verbs would lead to better generalization. Likewise, although computational models suggested that exposure to verbs whose semantic and phonological properties were not well aligned with the morpheme of interest would trigger generalization, testing this prediction in children was the necessary confirmatory step in order to apply this hypothesis to practice (Gupta, 2008). Observational studies that simply describe the patterns of use of children with language impairment would have confirmed that these children use language differently but do not address the learning process. Only a treatment study examining how the statistical regularities of English influence learning could provide us with this information indicating the critical role of the distribution of the input on the learning process.

Although further corroboration of the outcomes of our studies is necessary, our findings suggest that clinicians should consider avoiding the selection of verbs exclusively based on curricular themes or because they align with child interests and instead may want to consider the role that the verb's lexical and phonological properties may play in treatment. One approach would be to use the verb rankings reported in studies such as Owen Van Horne and Green Fager (2015) and Appendix A of Owen Van Horne et al. (2017) or to examine the factors described in Table 2 of this study as a way to select verb targets that are not a part of this subset. Other recent reports suggest that verb variability is another important

Table 6. Regression model for language samples examining the influence of condition assignment and time on past tense *-ed* production.

Random effects	Variance	SD		
Subject	0.56	0.75		
Story	0.01	0.11		
Fixed effects	β	SE	Z	p
Intercept				
Ref values: pretest, easy first	-0.53	0.29	-1.80	.07
Immediate post	0.32	0.21	1.51	.13
Delayed post	0.32	0.21	1.50	.13
Hard first	-0.08	0.40	-0.19	.85
Immediate post * HF	1.14	0.30	3.77	.0002**
Delayed post * HF	1.81	0.30	5.99	< .0001***

Note. $N = 1,437$; three stories, 20 subjects. HF = Hard First; AIC = Akaike Information Criterion, 1,697; BIC = Bayesian Information Criterion, 1,739.2; $df = 1,429$; LL = Log Likelihood, 840.5.

** $p < .001$. *** $p < .0001$.

factor to consider (Plante et al., 2014). Specifically, Plante and colleagues demonstrated that using 24 unique verbs within a treatment session focused on verb morphology leads to greater gains than using 12 unique verbs. In fact, it may be that a high-variability approach to treatment and a hard first approach to treatment are not that different. Selecting hard verbs for use in treatment likely increases the variability in the verbs that children hear, because everyday life is much more likely to consist of exposure to easy verbs. Providing exposure to a large number of different verbs likely also increases the children's exposure to verbs that are not well aligned with the targeted morpheme. Future work will be required to separate these two approaches.

We stated at the outset that the goal of this approach is to promote generalization. We succeeded in obtaining generalization across verbs and contexts. A key contribution of this article is showing that the gains reported in Owen Van Horne et al. (2017) are not restricted to the elicited production probes that were included as pre-post measures. This is important because there was always the possibility that practice effects during within-treatment probes might have influenced the overall outcomes at the two posttesting points. Here we show with language sample data that children were learning to use past tense *-ed* better in the hard first condition and that this is not restricted to artificially constructed elicited production tasks. In fact, although performance on the structured probes was approximately stable from immediate to delayed post-test, performance on the language sample showed continued growth. At first this seems counterintuitive, but it may suggest that ongoing consolidation and carryover occurred prior to transfer to more naturalistic activities. Given that our training took place over several weeks to months and that spacing varied across participants, our results do not map directly onto published accounts of memory and consolidation. Nonetheless, the passage of time may have led to more abstract representations of the regular past tense and thus to more widespread application of *-ed* to a variety of stems in spontaneous contexts. Further manipulations of spacing and timing may provide useful information about generalization over time and provide guidance to clinicians on when direct intervention can be stopped while gains continue without direct treatment.

Unfortunately, but unsurprisingly, we did not observe generalization across morphemes. This finding is well documented elsewhere (Leonard et al., 2004; Leonard, Camarata, Pawłowska, Brown, & Camarata, 2006, 2008; Plante et al., 2014). One might have expected that recognizing that past tense *-ed* could apply to atelic verbs would also lead to the recognition that third-person singular *-s* could be used with telic verbs; this is the prediction of Li and Shirai's (2001) work. In our study, the use of third-person singular *-s* also did not demonstrate growth. This may be due to two measurement-related issues. First, obligatory contexts for present tense were rarer than other morpheme types given the scoring rules we implemented for deciding when to code for present and past tense. Second,

the children started out near ceiling. Both small numbers of data points and measurement near ceiling weaken the inferences we can draw from the logistic regression models that we used to analyze our data. Similarly, we might have hoped that learning to use past tense *-ed* would lead to an understanding of pastness more broadly and carry over to irregular forms or the use of *was* and *were* as copula and auxiliary forms. Although other past tense forms increased in accuracy, this was observed in both groups and therefore cannot be separated from general maturation. Despite these measurement-related caveats, we conclude that cross-morpheme generalization is difficult to obtain. It is important to realize that children learn what we treat, and we will have to systematically treat grammatical markers if we want children to be competent users of these markers. To fully test Li and Shirai's model, we would need to expose children to misaligned verb + inflection combinations of all morphemes. We did not do this because children with DLD seem to learn any individual morpheme less well when they are exposed to multiple morphemes at the same time (Leonard et al., 2004, 2006, 2008; Plante et al., 2014). Further work that examines how to successfully combine treatment for multiple morphemes while maximizing generalization would be beneficial.

In conclusion, we have demonstrated that the children in the hard first condition, as a group, made greater gains in accuracy than the children in the easy first condition. We made this hypothesis on the basis of a computational model that provided visualizations of complex statistical information present in everyday use of English. The results of this study indicate that selecting verbs for training children to use past tense *-ed* based on their lexical, aspectual, and phonological properties can influence treatment outcomes. "Easy" verbs may promote early success, but "hard" verbs lead to greater generalization. Clinicians should consider how the properties of the words they use in treatment interact with the grammatical targets being treated.

Acknowledgments

This research was funded by National Institute on Deafness and Other Communication Disorders Grant K23-DC 013291 awarded to Amanda J. Owen Van Horne while she was at the University of Iowa. Stimuli development and pilot work was made possible through a grant from the ASH Foundation, awarded to Amanda J. Owen Van Horne. Lauren Seemann and Diane Buffo coordinated the intervention study and data management. Members of the Grammar Acquisition Lab at the University of Iowa assisted with intervention provision, data collection, transcription, and analysis. School districts across Iowa supported the recruitment process. We would especially like to thank the assistance provided by 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

- Alt, M., Meyers, C., & Ancharski, A. (2012). Using principles of learning to inform language therapy design for children with specific language impairment. *International Journal of Language & Communication Disorders*, 47(5), 487–498.
- Baayen, R. H. (2008). *Analyzing linguistic data*. Cambridge, United Kingdom: Cambridge University Press.
- Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2–3), 150–177. <https://doi.org/10.1080/00437956.1958.11659661>
- Bishop, D. V. M., Snowling, M., Thompson, P. A., Greenhalgh, T., & the CATALISE Consortium. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLoS One*, 11(7), e0158753. <https://doi.org/10.1371/journal.pone.0158753>. s006
- Bishop, D. V. M., Snowling, M., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 58(10), 1068–1080. <https://doi.org/10.1111/jcpp.12721>
- 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(1), 281–294. [https://doi.org/10.1044/1092-4388\(2012/11-0112\)](https://doi.org/10.1044/1092-4388(2012/11-0112))
- Borovsky, A., & Elman, J. L. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language*, 33(4), 759–790. <https://doi.org/10.1017/S0305000906007574>
- 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(2), 237–255. https://doi.org/10.1044/2015_AJSLP-14-0105
- Comrie, B. (1976). *Aspect*. New York, NY: Cambridge University Press.
- Crystal, D. (1985). Some early problems with verbs. *Child Language Teaching and Therapy*, 1, 46–53.
- Dawson, J. I., Stout, C. E., & Eyer, J. A. (2003). *SPELT-3: Structured Photographic Expressive Language Test—Third Edition*. Dekalb, IL: Janelle Publications.
- Dunn, L., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test—III*. San Antonio, TX: Pearson.
- Eisenberg, S. L. (2013). Grammar intervention: Content and procedures for facilitating children's language development. *Topics in Language Disorders*, 33(2), 165–178. <https://doi.org/10.1097/TLD.0b013e31828ef28e>
- Fey, M. E., Cleave, P. L., Long, S. H., & Hughes, D. L. (1993). Two approaches to the facilitation of grammar in children with language impairment: An experimental evaluation. *Journal of Speech and Hearing Research*, 36(1), 141–157.
- Finestack, L. H. (2014). Language learning of children with typical development using a deductive metalinguistic procedure. *Journal of Speech, Language, and Hearing Research*, 57(2), 509–523.
- Finestack, L. H., & Fey, M. E. (2009). Evaluation of a deductive procedure to teach grammatical inflections to children with language impairment. *American Journal of Speech-Language Pathology*, 18(3), 289–302.
- Gierut, J. A. (2001). Complexity in phonological treatment: Clinical factors. *Language, Speech, and Hearing Services in Schools*, 32(4), 229–241.
- Goldman, R., Fristoe, M., & Williams, K. T. (2000). *Goldman-Fristoe Test of Articulation—Second Edition*. San Antonio, TX: Pearson.
- Gupta, P. (2008). The role of computational models in investigating typical and pathological behaviors. *Seminars in Speech and Language*, 29(3), 211–225. <https://doi.org/10.1055/s-0028-1082885>
- Hassink, J. M., & Leonard, L. B. (2010). Within-treatment factors as predictors of outcomes following conversational recasting. *American Journal of Speech-Language Pathology*, 19(3), 213–224. [https://doi.org/10.1044/1058-0360\(2010/09-0083\)](https://doi.org/10.1044/1058-0360(2010/09-0083))
- Hills, T. T. (2013). The company that words keep: Comparing the statistical structure of child- versus adult-directed language. *Journal of Child Language*, 40(3), 586–604. <https://doi.org/10.1126/science.274.5294.1926>
- Johnson, B. W., & Fey, M. E. (2006). Interaction of lexical and grammatical aspect in toddlers' language. *Journal of Child Language*, 33(2), 419–435. <https://doi.org/10.1017/S0305000906007410>
- Kaufman, A. S., & Kaufman, N. L. (2004). *KBIT-2: Kaufman Brief Intelligence Test—Second Edition*. San Antonio, TX: Pearson.
- Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research*, 46(4), 773–787.
- Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Berlin, Germany: Springer.
- Leonard, L. B. (1975). Modeling as a clinical procedure in language training. *Language, Speech, and Hearing Services in Schools*, 6(2), 72–85.
- Leonard, L. B., Camarata, S. M., Brown, B., & Camarata, M. N. (2004). Tense and agreement in the speech of children with specific language impairment: Patterns of generalization through intervention. *Journal of Speech, Language, and Hearing Research*, 47(6), 1363–1379.
- Leonard, L. B., Camarata, S. M., Pawłowska, M., Brown, B., & Camarata, M. N. (2006). Tense and agreement morphemes in the speech of children with specific language impairment during intervention: Phase 2. *Journal of Speech, Language, and Hearing Research*, 49(4), 749–770.
- Leonard, L. B., Camarata, S. M., Pawłowska, M., Brown, B., & Camarata, M. N. (2008). The acquisition of tense and agreement morphemes by children with specific language impairment during intervention: Phase 3. *Journal of Speech, Language, and Hearing Research*, 51(1), 120–125.
- 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. *Child Language Teaching and Therapy*, 21(10), 747–758. <https://doi.org/10.1080/02699200701495473>
- Leonard, L. B., Deevy, P., Kurtz, R., Finneran, D., Owen, A. J., Polite, E., & Elam, D. (2007). Lexical aspect and the use of verb morphology by children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 50(3), 759–777. [https://doi.org/10.1044/1092-4388\(2007/053\)](https://doi.org/10.1044/1092-4388(2007/053))
- Li, P., & Shirai, Y. (2001). *The acquisition of lexical and grammatical aspect*. Berlin, Germany: Mouton De Gruyter.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk, Vol 2: The database* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Marchman, V. A. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science: A Multidisciplinary Journal*, 21(3), 283–303.

- Marchman, V. A., Wulfeck, B., & Weismer, S. E. (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(1), 206–219.
- Mayer, M. (1970). *Frog, where are you?* New York, NY: Dial Books.
- Mayer, M. (1975). *One frog too many*. New York, NY: Dial Books.
- Mayer, M., & Mayer, M. (1969). *A boy, a dog, a frog, and a friend*. New York, NY: Dial Books.
- Miller, J. F., & Iglesias, A. (2012). *Systematic Analysis of Language Transcripts: SALT, Research Version 2012* [Computer program]. Madison, WI: SALT Software.
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130299. <https://doi.org/10.1016/j.cognition.2011.10.004>
- 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(1), 62–74.
- Owen Van Horne, A. J., Fey, M., & Curran, M. (2017). Do the hard things first: A randomized controlled trial testing the effects of exemplar selection on generalization following therapy for grammatical morphology. *Journal of Speech, Language, and Hearing Research*, 60(9), 2569–2588. https://doi.org/10.1044/2017_JSLHR-L-17-0001
- 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(6), 605–616. <https://doi.org/10.1111/j.1540-4781.2009.00895.x>
- 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(2), 103–115.
- Pinker, S. (2000). *Words and rules: The ingredients of language*. New York, NY: Harper Collins.
- 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(4), 530–545. https://doi.org/10.1044/2014_AJSLP-13-0038
- Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition*, 38(1), 43–102.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77(3), 353–363.
- Redmond, S., & Rice, M. L. (2001). Detection of irregular verb violations by children with and without SLI. *Journal of Speech and Hearing Disorders*, 44, 655–669.
- Riches, N. (2015). Past tense –ed omissions by children with specific language impairment: The role of sonority and phonotactics. *Clinical Linguistics & Phonetics*, 29(6), 482–497.
- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press.
- Rvachew, S. (2005). Stimulability and treatment success. *Topics in Language Disorders*, 25, 207–219.
- Rvachew, S., & Nowak, M. (2001). The effect of target-selection strategy on phonological learning. *Journal of Speech, Language, and Hearing Research*, 44(3), 610–623.
- Shirai, Y. (2015). Frequency effects in grammatical development: A cross-linguistic, functional approach to form-function mapping. *Journal of Child Language*, 42(2), 312–315. <https://doi.org/10.1017/S030500091400083X>
- Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition*, 14(1), 17–26.
- Thompson, C., & Shapiro, L. P. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology*, 16(1), 30–42. [https://doi.org/10.1044/1058-0360\(2007\)005](https://doi.org/10.1044/1058-0360(2007)005)
- 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(4), 516–528. <https://doi.org/10.1017/CBO9780511627705.016>
- Warmington, M., Hitch, G., & Gathercole, S. E. (2013). Improving word learning in children using an errorless technique. *Journal of Experimental Child Psychology*, 114(3), 456–465. <https://doi.org/10.1016/j.jecp.2012.10.007>
- Weiler, B. (2013). Verb selection and past-tense morphology. *Topics in Language Disorders*, 33(2), 152–164. <https://doi.org/10.1097/TLD.0b013e31828f50b8>
- Wulff, S., Ellis, N. C., Roemer, U., Bardovi-Harlig, K., & Leblanc, C. (2009). The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal*, 93(3), 354–369.