

## Review Article

# Learning Without Trying: The Clinical Relevance of Statistical Learning

Elena Plante<sup>a</sup> and Rebecca L. Gómez<sup>a</sup>

**Purpose:** Statistical learning research seeks to identify the means by which learners, with little perceived effort, acquire the complexities of language. In the past 50 years, numerous studies have uncovered powerful learning mechanisms that allow for learning within minutes of exposure to novel language input.

**Method:** We consider the value of information from statistical learning studies that show potential for making treatment of language disorders faster and more effective.

**Results:** Available studies include experimental research that demonstrates the conditions under which rapid

learning is possible, research showing that these findings apply to individuals with disorders, and translational work that has applied learning principles in treatment and educational contexts. In addition, recent research on memory formation has implications for treatment of language deficits.

**Conclusion:** The statistical learning literature offers principles for learning that can improve clinical outcomes for children with language impairment. There is potential for further applications of this basic research that is yet unexplored.

Over time, the professional debates on how to structure language treatment have swung between the value of highly structured, drill-like treatments versus treatments that mimic naturalistic interactions or that follow the child's lead. We leave these debates aside to address a more pressing and practical problem. Treatment in its current forms simply takes too long. Even with substantial time and effort, the amount of change achieved can be disheartening. In their systematic reviews, Law and colleagues reported that, although better than no treatment, language intervention studies produce modest results overall, with longer treatments generally providing the best results (Law, Garrett, & Nye, 2004, 2010). In addition, generalization of learning has been largely unaddressed, despite the frequent concern that children with developmental language disorder often seem to leave their new skills in the treatment room (Haley, Hulme, Bowyer-Crane, Snowling, & Fricke, 2017; Kamhi, 1988, 2014). We suggest that one important reason for this situation is that treatment research to date has had relatively little contact with basic research

on how rapid, generalizable language learning occurs. In this review article, we provide a brief overview of the nature of statistical learning processes. We then propose five principles derived from statistical learning studies of typically developing learners that we believe offer promise for improving treatment outcomes for learners with developmental language disorders.

## Implicit “Statistical” Learning

Implicit learning is a process in which learners extract regularities from the world around them without conscious intent or knowledge of these patterns. Such learning contrasts with explicit teaching on the part of adults (e.g., “Wheat is a plant because it grows in the ground”; “When there are two, we say /s/”) or attempts by the learner to think explicitly about what constitutes correct language use (e.g., Should I use “he” vs. “him” this time?). Instead, implicit learning capitalizes on the learner's own cognitive biases for tracking structure in the input. Moreover, for typically developing learners, there is evidence that statistical patterns, once learned, are retained over time (Frank, Tenenbaum, & Gibson, 2013; Perry, Samuelson, Malloy, & Schiffer, 2010).

Although those with language disorders often do not appear to detect statistical structure as readily as typically developing learners (Evans, Saffran, & Robe-Torres, 2009; Plante, Gómez, & Gerken, 2002; Richardson, Harris, Plante, & Gerken, 2006), we have evidence that we can enhance

<sup>a</sup>The University of Arizona, Tucson

Correspondence to Elena Plante: [eplante@email.arizona.edu](mailto:eplante@email.arizona.edu)

Editor-in-Chief: Mary Alt

Editor: Shelley Gray

Received November 20, 2017

Revision received January 18, 2018

Accepted February 7, 2018

[https://doi.org/10.1044/2018\\_LSHSS-STLT1-17-0131](https://doi.org/10.1044/2018_LSHSS-STLT1-17-0131)

**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.

and sometimes even normalize learning for those with developmental language disorder when we employ methods that enhance the salience of statistical structure (Aguilar & Plante, 2014; Grunow, Spaulding, Gómez, & Plante, 2006; Plante, Vance, Moody, & Gerken, 2013; Torkildsen, Dailey, Aguilar, Gómez, & Plante, 2013). This also appears to be true in language treatment studies that use methods grounded in statistical learning theory (Aguilar, Plante, & Sandoval, 2018; Alt, Meyers, Oglivie, Nicholas, & Arizmendi, 2014; Meyers-Denman, & Plante, 2016; Plante, Tucci, Nicholas, Arizmendi, & Vance, 2018). We note here that learning can occur without the benefit of applying principles of statistical learning. However, a statistical learning perspective predicts that learning will be more efficient and effective when statistical learning principles inform the clinician's input to children during treatment.

The idea of an implicit means for language learning began as early as the 1960s as a then tentative contrast to the nativist view of language acquisition. In an early experimental approach, Reber (1967, 1969) demonstrated that adults could learn an artificial grammar represented by alphabet letters simply by viewing grammatical letter strings presented one at a time on flashcards. Despite being able to learn the underlying rules of the grammar, the participants were largely unaware of specific rules in the examples provided. Nor could they explicitly and accurately convey the basis of their grammaticality judgments in words. Since then, language researchers have applied similar experimental approaches to many forms of learning with language-like stimuli (e.g., Morgan, Meier, & Newport, 1987, 1989; Morgan & Newport, 1981; Reber, 1989; Valian & Coulson, 1988) as well as in multiple sensory modalities (Fiser & Aslin, 2002a, 2002b; Kirkham, Slemmer, & Johnson, 2002; Saffran, Johnson, Aslin, & Newport, 1999). Accordingly, statistical learning is considered a general purpose learning mechanism, rather than one that is specific to language acquisition.

Two lines of research are particularly informative for our purpose here. One focuses on operations underpinning detection of sequential structure; the other focuses on learning of morphosyntax that requires distributional learning of cues to category structure. Each of these is relevant to different aspects of language learning and therefore treatment. Importantly, there is evidence that learners with developmental language disorder can learn both types of structures.

### ***Learning Sequential Structure***

One of the earliest challenges in the course of language acquisition lies in the detection of sequential structure. Sequential probabilities express the likelihood of one element (e.g., a letter, a sound, a syllable, a word) predicting the occurrence of another. This is how learners recognize a sequence of sounds or syllables as representing a word. Consider an English language example: the phrase *pretty baby*. The transition between the two syllables of *pretty* has a higher statistical likelihood than the transition between *ty* and *ba*. The syllable *pre* predicts *ty* with greater frequency

than *ty* predicts *ba*. In a seminal review article, Saffran, Newport, and Aslin (1996) showed that typically developing 8-month-olds can track these types of sequential transitional probabilities and differentiate between nonword syllable combinations that represent word-like sequences and random syllable co-occurrences. The ability to segment continuous speech on the basis of syllable-level transitional probabilities gave rise to the term “statistical” learning. This use of statistical information does not end with word segmentation. Typically developing infants readily map segmented word forms onto novel word referents (Graf Estes, Evans, Alibali, & Saffran, 2007). This subsequent word meaning mapping capitalizes on the co-occurrence statistics for the word form and a real-world referent.

### ***Learning Morphosyntactic Cues to Category Structure***

Category structure characterizes multiple aspects of language. Words and their meanings can be organized into supraordinate and subordinate categories (e.g., animals–mammals–rodents). Words also have syntactic categories (e.g., articles, nouns, verbs, prepositions) that affect how they can be ordered within a grammatical utterance. Therefore, learning category structure is critical for a semantically rich lexicon and an internal representation of syntax that can be used to generate new utterances.

In an early demonstration of category learning, Braine (1987) demonstrated that typically developing learners quickly form morphosyntactic categories through statistical learning. His artificial grammar paralleled existing English categories and “rules” for their combinations. In English, the members of the word category of *noun* often follow the occurrence of an article (e.g., *a* or *the*). Verbs can follow an auxiliary (e.g., *is* or *are*). These combinations produce phrase structures such as “*the boy*” and “*a car*,” or “*is kicking*” and “*are kicking*,” but not “*a kicking*” or “*is car*.” The predictable relation between categories of elements promotes a distributional analysis of statistical cues by typically developing learners that allows learners to recognize and generalize the underlying syntactic “rules.” Typically developing learners can use a large variety of cues to category membership (Frigo & McDonald, 1998; Gerken, Wilson, & Lewis, 2005; Gómez & Lakusta, 2004; Hall, Owen Van Horne, & Farmer, 2018; Reeder, Newport, & Aslin, 2013; Valian & Coulson, 1988), including phonetic content, semantic content, co-occurrence statistics, and the relative position of words within a string. Individuals with developmental language disorders likewise show category formation through statistical learning that can be similar to their healthy peers (Hall, Owen Van Horne, McGregor, & Farmer, 2017; Torkildsen et al., 2013). This indicates that statistical learning approaches have the potential to support this type of learning in treatment.

### ***The Neurological Reality of Statistical Learning***

Neuroimaging evidence indicates that statistical learning is fundamentally different from nonstatistical learning.

Plante et al. (2015) asked adults to listen to full-length Norwegian sentences (a language unknown to the learners). Half the adults listened to sentences that provided strong statistical cues to aid identification of adjacent syllable pairs that represented real Norwegian words. The other half heard nearly identical input, but the statistical cues that identified these words were removed. Adults with access to statistical cues were much more proficient in identifying words than learners denied these cues. This reinforced the notion that statistical learning results in rapid language acquisition. Importantly, statistical learning activated a robust learning network that involved the integration of multiple cortical regions (see Figure 1). In comparison, learners prevented from learning through statistical means showed a much more spatially limited and weakly activated learning network. This offers neurological confirmation that the rapid acquisition associated with statistical learning recruits different neural resources than nonstatistical learning. A subsequent study showed that adults with and without developmental language disorders activated a common neural network during the word segmentation task, indicating both groups were similarly engaged in statistical learning. Adults with developmental language disorder seemed to require more neural effort to perform the same task. Despite being as successful as their healthy peers in segmenting words, those with developmental language disorders displayed higher activation levels than their healthy peers (Plante, Patterson, Sandoval, Vance, & Asbjørnsen, 2017), consistent with their impaired status and the idea that less efficient processing requires more neural effort. Importantly, however, those with developmental language disorders

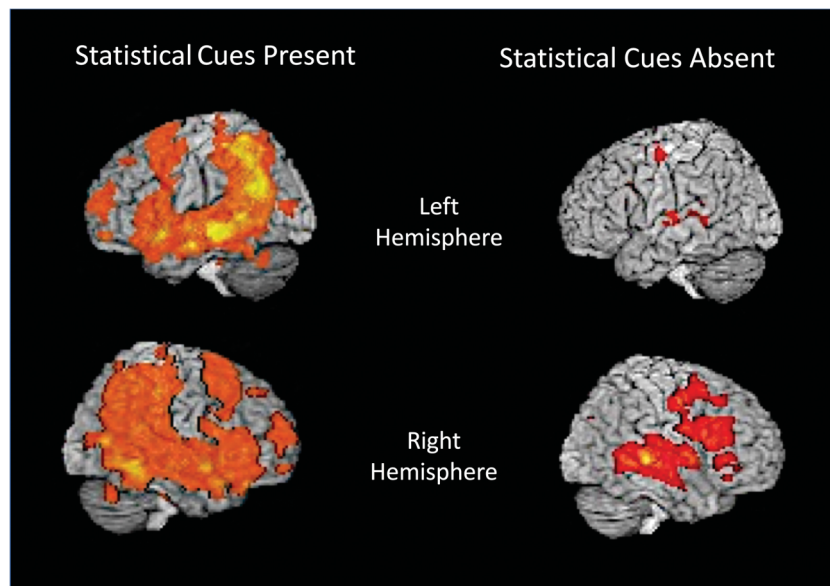
used the statistical learning network when presented with statistically structured input.

### *Characteristics of Statistical Learning Studies*

In looking toward adaptation of statistical learning to treatment contexts, it is worth noting several features common to these statistical learning studies. First, in each of the research studies described, learners simply listened to recorded input for several minutes and were later tested on what they had learned. Language input was not provided in naturalistic contexts or even through interpersonal interactions. Second, these studies isolate a particular statistical cue so that new linguistic forms can be learned rapidly and often use nonwords and artificial grammars to achieve this control. It is important to note, however, that rapid implicit learning has been documented with natural language stimuli as well (e.g., Eidsvåg, Austad, Plante, & Asbjørnsen, 2015; Gerken et al., 2005; Kittleson, Aguilar, Tokerud, Plante, & Asbjørnsen, 2010; Pelucchi, Hay, & Saffran, 2009a, 2009b; Plante et al., 2015; Sandoval, Patterson, Dai, Vance, & Plante, 2017). Third, learners were never provided any information concerning what to listen for. Instead, detection of statistical structure by the learner is enhanced by the careful experimenter control over the nature of the input in each study. We propose that careful clinician control over the nature of the input in treatment can similarly facilitate learning.

Finally, statistical learning paradigms do not provide feedback to the learner. Feedback-based learning is known to activate different neural regions than statistical learning

**Figure 1.** Neuroimaging data showing the difference in learning networks when strong statistical cues to language structure are available in the input versus when they are not. Colors indicate regions of activation. Identical colors indicate subregions of the learning network that show the same activation timing as revealed by independent component analysis. Data from Plante et al. (2015).



(Opitz, Ferdinand, & Mecklinger, 2011). This suggests that the presence of feedback shifts learners away from using just those cognitive processes that promote statistical learning. Furthermore, there is some question concerning whether those with a developmental language impairment can make use of the feedback provided to them. Arbel and Donchin (2014) have demonstrated that not only do children with developmental language disorder show poor feedback-related learning compared to their typically developing peers but they also do not show the normal brain response to feedback. This evidence suggests that children with developmental language disorder do not process feedback in the same way and are less able to take advantage of feedback for shaping learning as well as their typically developing peers. Given this, a learning method that does not require incorporating feedback might be particularly attractive for those with developmental language disorder.

The following sections of this article will explore specific strategies for making statistical information salient to the learner in treatment contexts. We describe five principles derived from statistical learning research that are applicable to the treatment of children and adults with developmental language disorder. An additional principle, concerning input complexity, is described elsewhere (Alt, Meyers, & Ancharski, 2012; Van Horn, Curran, Larson, & Fey, 2018).

## The Regularity Principle, Encompassing Frequency of Occurrence and Consistency

Learners seek regularity in the input they receive (the Regularity Principle). In the world at large, variability abounds. There are hundreds of tree species. There are thousands of flower varieties. Yet none of us mistake a mature tree for a mature flower. This is because we integrate the features that regularly occur for each class of plant and form categories for all things tree and all things flower. The regularities common to both of these categories allow us to form the supraordinate category of “plant.” In this way, we are able to develop the semantic webs that give depth to the meaning of the words we know. For typically developing children acquiring language, it is not necessary to explain these similarities and differences. Mere exposure to multiple examples of category members and the category label is sufficient to induce the formation of a word class (Perry et al., 2010; Twomey, Ranson, & Horst, 2014; Vlach & Sandhofer, 2011).

### Frequency of Occurrence

To translate the Regularity Principle into clinical practice, two factors must be considered. The first is frequency of occurrence. The typical statistical learning study presents as many informative examples as possible within short periods of time, resulting in high-frequency and high-density input. For typically developing learners, this leads to rapid learning (i.e., within a single session). In contrast, children with developmental language disorder can require twice as much

input as their typically developing peers to identify individual words in the Saffran et al. (1996) nonword segmentation task (Evans et al., 2009). Children with developmental language disorder also require more exposures than their typically developing peers to acquire new words and their meanings in vocabulary-learning studies (e.g., Gray, 2004; Rice, Oetting, Marquis, Bode, & Pae, 1994). In the domain of morphosyntax, children with developmental language disorder require therapeutic treatment doses at higher density (more exposures per unit of time) than these events occur in the real world for morpheme learning to occur (Fey, Krulik, Loeb, & Proctor-Williams, 1999). Therefore, part of applying the Regularity Principle is to assure that enough examples of the learning target (e.g., new lexical labels, object-label pairings, grammatical elements) occur with a high enough frequency within the treatment period to be effective. Although what constitutes “high enough” deserves greater research attention, children with developmental language disorders benefit from hearing many more examples of each therapy target within a treatment session than what typically developing children experience in their natural environment.

Consider the following example, modeled after a treatment study by Alt et al. (2014). This team sought to facilitate vocabulary learning in late-talking toddlers. They presented input similar to that found in Table 1 to 2-year-olds who began treatment between the fifth and 15th percentile for words known on the MacArthur-Bates Communicative Development Inventories-Second Edition (Fenson, Marchman, Thal, Dale, Reznick, & Bates, 2006). The words targeted for each child to learn were presented at rates ranging from 4.85 to 14.67 per minute. At these rates, each target word occurred at very high frequencies and densities within each session. Under these conditions, children produced words targeted in treatment at much higher rates than they learned untreated words.

### Consistency

The second parameter under the Regularity Principle involves making the treatment target the most consistent event the child encounters during a therapy session. Consider again the input in Table 1. The target word in this example

**Table 1.** Examples of input provided to facilitate vocabulary learning based on Alt et al. (2014).

---

Here is your shoe.  
Look, one shoe is blue.  
Shoes go on feet.  
This is a small shoe.  
Dolly's shoe is here.  
He needs a shoe.  
I found a shoe in here.  
Shoes are for walking.  
This shoe doesn't fit.  
I'd like a shoe...

---

*Note.* Sixty-four sentences containing target words were presented, during interactive play, at rates ranging from 4.85 to 14.67 per minute.

---



(shoe) not only occurs very frequently in the input but it is the only word that appears consistently across sentences. Indeed, the consistency of “shoe” is so much higher than for all other words; viewers of this table are likely to surmise the target of learning without being told explicitly. This type of consistency for training targets benefits learning in treatment contexts.

High frequency and consistency can also be applied to treatment of morphological errors. Table 2 presents possible types of input a clinician might present to treat omissions of the third-person *-s* agreement verb marker. In this hypothetical scenario, three clinicians use farm animals as a context for treatment. In Table 2, Clinician 1 provides highly consistent verbal input, with the pronoun “he” and the verb “stands” in each utterance. Because the child is sensitive to elements that occur regularly, he or she will likely form the idea that both “he” and “stands” and their co-occurrence are all important. Accordingly, when tested on his or her learning, the child will likely produce “He stands” (sometimes regardless of the gender of the actor). Because the input offered only evidence for “He stands” as good utterances, there will be no generalization beyond this input. Indeed, because of the regularity and co-occurrence of “he” and “stands,” there is a danger of the child encoding “He stands” as a single unit, rather than two independent words. In this scenario, regularity of multiple elements promotes the wrong interpretation about the nature of the input.

Clinician 2 provides a little more varied input in that the root verb varies, but “he” and the verb morpheme “-s” still appear regularly in the input. Unfortunately, this often produces generalization of the pronoun “he” but not of the “-s” morpheme. There are several possible reasons for this. First, this input meets the frequency criterion for the Regularity Principle, in that *-s* is highly frequent. However, “he” is just as frequent, such that no one element is the most regularly occurring. “He” also gets an additional saliency boost because the /i/ in “he” has a higher acoustic intensity

than the /s/ in “-s” (cf. Rom & Leonard, 1990). In addition, “he” always occurs first, so that it gets a primacy boost for memory encoding because initial items (and final items) are remembered better than middle items in a string (Endress, Scholl, & Mehler, 2005; Reber & Allen, 1978). Clinician 3 solves this problem by making the *-s* morpheme both frequent in the input and the most regular element heard. The number of times any other morpheme, bound or free, occurs pales in comparison to the regular occurrence of “-s.” In this case, “-s” is the element that generalizes to untrained verbs and new sentence types. This reflects optimal frequency and consistency in use of the Regularity Principle in treatment.

One word of caution about application of the Regularity Principle to treatment is warranted. Capitalizing on regularity requires the child to track specific lexical or grammatical forms in the input. This ability does not necessarily apply to entire linguistic classes. For example, attempting to train the class of pronouns (e.g., she, he, they, her, him, them) or verb forms (e.g., auxiliary are, auxiliary is) can prevent optimization of the parameters of the Regularity Principle within an individual session. First, presenting multiple pronouns makes each pronoun less frequent in the input, and at some frequency level, individual pronouns may become too infrequent to be regularly occurring. Second, when multiple pronouns co-occur in the clinician’s input, none will stand out as the most regular aspect of that input. Under these conditions, learning is not likely to extend to untreated errors. In Table 2, the child treated by Clinician 3 generalizes the third-person *-s* but substitutes “her” for “she,” reflecting a separate pronoun error (see column 4, row 6).

Yet unknown is whether learning a single exemplar of a grammatical class (e.g., auxiliary is) speeds learning of other members of the grammatical class (e.g., auxiliary are). There is reason to think that it would, from the experimental statistical learning literature. If typically developing learners master fundamental operations in the learning of category structure as described above, then this previous learning

**Table 2.** The effect of different statistical patterns of input to children during treatment of third-person singular use in terms of generalization to untreated examples.

Treatment events	Clinician 1	Clinician 2	Clinician 3
Clinician input	He [cow] stands. He [sheep] stands. He [donkey] stands. . . .	He [cow] chews grass. He [sheep] stands. He [donkey] kicks. . . .	The cow chews grass. One sheep stands. He [donkey] kicks. . . .
	He [duck] stands.	He [duck] swims over.	A duck swims over.
Child forms generalizations (legal and illegal)	He [horse] stands. He [girl] stands.	He [horse] come. He [girl] look.	He [horse] runs. Her looks. The boy runs. Baby sleeps.
Child does NOT generalize to...	The boy runs. Baby sleeps.	He runs. Baby sleeps.	

*Note.* The total number of doses from each clinician is 24 recasts containing the target morpheme. Input examples are purposely abbreviated for space considerations.

can boost new learning of related structures (Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007). For example, knowledge gained about sentence structure in an artificial language transfers to recognition of that same structure when used with words or word combinations never previously heard (Gerken, 2004; Gerken et al., 2005; Gómez & Gerken, 1999; Gómez & Lakusta, 2004; Marcus, Vijayan, Bandi Rao, & Vishton, 1999). In this way, we might expect that training children on an “is + verb + ing” structure would lead to even faster subsequent learning of an “are + verb + ing” structure. If this basic research translates to treatment contexts, then it may prove more efficient to train a single structure at a time than to try to train multiple linguistically related targets simultaneously.

### The Variability Principle: Strategic Variation of Nontarget Elements

The variability principle expands on the Regularity Principle by dictating the form of nontarget words provided in the input. Although elaborated here, it has been discussed elsewhere with reference to child language treatment (Alt et al., 2012; Plante et al., 2014). The variability principle states that high input variability for the nontarget elements promotes learning of the treatment target. As much as the target of learning must occur regularly in the input, the strategic use of variability in nontarget items serves to make the highly regular target even more salient. Learners appear to track what is regular in the input and come to ignore what is variable. We see the purposeful and strategic use of variability in the input listed in Table 1, and Table 2 for Clinician 3. In both cases, not only is the learning target the most regularly occurring element (shoe in Table 1; third-person *-s* morpheme in Table 2), but all other words used in the input vary as well.

The variability principle was first demonstrated by Gómez (2002). This work showed that variability in the nontarget elements could assist learning of a syntax-like form. Paired studies of typically developing infants and adults demonstrated that learners could acquire relatively difficult-to-learn nonadjacent dependencies in an artificial grammar, but only under certain input conditions. The nonadjacent dependency used in this study has an English equivalent in the “is verbing” grammatical structure (e.g., *is* predicts *ing*, with any number of verbs coming between). When learners heard 12 or fewer examples of the middle element in the grammar, no learning occurred. However, when the variability of the middle element increased to 24 examples, learning occurred rapidly for both infants and adults. This was true even though the total number of examples provided to learners was identical across variability groups (e.g., those who heard only three unique examples heard each example eight times, for a total of 24 presentations).

The variability principle has been translated successfully to developmental language disorder in experimental studies (Grunow et al., 2006; Torkildsen et al., 2013), showing that even those with developmental language

disorder benefit from high input variability. In these studies, presentation of 24 unique examples of the target grammar to learners, without any instruction concerning what to listen for, led to learning and generalization of that grammar to untrained exemplars. In contrast, those learners hearing 12 exemplars, each presented twice, showed little or no learning. These experimental findings were translated to treatment studies for children with developmental language disorder (Aguilar et al., 2018; Plante et al., 2014). In Plante et al. (2014), a variety of morphemes were targeted for treatment (one per child, following the Regularity Principle). Children who heard their target morpheme in 24 different verb contexts outperformed children who heard their target in 12-verb contexts, even though the lower variability group heard morpheme + verb pairings twice each for a total of 24 models.

In the domain of semantics, two studies capitalize on the variability principle for facilitating vocabulary learning. Alt and colleagues (2014), in the study discussed above, used highly regular target vocabulary words embedded into highly variable sentence frames. Activities and visuals corresponding to target words (e.g., shoe, dog, car, block) also varied within and across sessions. Aguilar et al. (2018) specifically examined whether varying object exemplars facilitated the formation of semantic classes for new words. Aguilar et al. (2017) presented previously unknown nouns to two groups of children who either saw three diverse examples of the object named (e.g., three hinges of different materials, sizes, and styles) or three identical examples (e.g., three identical hinges). Word labels for these objects were also presented in highly varied sentence contexts. Aguilar and colleagues then tested children to determine if they could correctly identify untrained examples of each object (a within-class generalization). Although these authors found no difference in learning during the 3-day training period, they did find a significant advantage for high-variability training in terms of the number of words retained approximately 3 weeks later. Children in the high-variability condition maintained their new words without additional training, whereas children in the low-variability condition showed a pattern of forgetting.

The variability principle does not just facilitate oral language skills. There are also examples of how strategic manipulation of nontarget input can assist in spelling (Apfelbaum, Hazeltine, & McMurray, 2013) and mathematics (Powell, Driver, & Julian, 2015). In the case of spelling, Apfelbaum et al. (2013) addressed training sound-symbol correspondence for vowels in English spellings. Mapping of vowels to their corresponding letters is more difficult than consonants for spelling for several reasons. First, vowels lack articulatory contact cues that consonants frequently provide. Second, vowels often occur in word medial positions, and there are one-to-many mappings for vowel symbols and sounds in English. Despite these challenges, when children were provided training with letter frames in which both initial and final consonants varied (e.g., *cat*, *ran*, *tag*), learning was superior to when they learned vowel sounds in word frames such as “cat, hat,

sat, ...” in which only the initial consonant varied. The variability advantage occurred whether the vowel sound was a single short vowel or a diphthong. Note also that the most regular element in the low-variability example was the vowel + final consonant rather than just the vowel. It may be that children were unable to generalize the vowels in this context because the combination “at” always occurred as a highly regular unit. This inadvertent regularity provides no evidence that the /æ/ vowel can occur in consonant contexts other than “at” or that the grapheme “a” is pronounced the same way no matter what consonant follows it. This promotes learning “at” as a unit, rather than learning the sound–symbol pairing represented by the vowel alone. Therefore, adding strategic variability to the consonant context provided evidence that the sound–symbol correspondence to be learned from the worksheet input was constant regardless of the initial or final consonants.

The variability principle emphasizes need for high input variability in the nontarget elements to facilitate learning. The data indicate that the converse is also true; high repetition of input can actually inhibit learning. It is easy for practitioners to assume that frequent repetition of a small sample of training items will be more helpful for children with limitations in language than a large amount of linguistically diverse input. Indeed, we frequently hear concerns that providing many, varied examples will “overwhelm” children with language disorders. In some sense, that is actually both true and beneficial. High-variability input prevents learners from trying to track each individual sound and syllable in the input because there are simply too many to hold in memory. Instead, highly variable input appears to encourage those with developmental language disorders to abandon attempts to remember exactly what they heard in favor of tracking the most regular aspect of the input (Plante et al., 2013). In this way, variation around the target increases target salience. It is better that every utterance provides a unique linguistic context for the training target than to repeat the same input even once (Gómez, 2002; Grunow et al., 2006; Plante et al., 2014; Torkildsen et al., 2013).

Although research has not yet defined the parameters precisely, the available research strongly suggests a minimum number of high-variability examples for learning to occur. Learners with developmental language disorder learn poorly when only 12 exemplars are provided and better with 24 exemplars (Grunow et al., 2006; Plante et al., 2014; Torkildsen et al., 2013), with the bare minimum likely between 12 and 24. The optimal number could be greater than 24. For this reason, the number of treatment examples for morphosyntax treatment would include at least 24 language “doses” or episodes of informative input provided to the child. For the example in Table 2, this would correspond to 24 clinician recasts to the child that contain the target morpheme. What might seem like a lot of variability (e.g., 12 exemplars) is actually insufficient for some types of morphosyntactic learning. That said, the minimum amount of variability necessary for learning can vary by linguistic target. For example, it appears fewer unique pairings of labels and referents are

needed for training word meanings (Aguilar et al., 2017; Perry et al., 2010). We know as few as three different physical exemplars are sufficient to induce learning of noun meanings that generalize to other objects of the same semantic category. Given the limited information we have on variability across linguistic targets, it is probably best to incorporate as much variability as possible into treatment methods, until more specific information is available.

## **Input Principle I: All Input Is Input When Learning Is Implicit**

Natural language input is noisy. All children are exposed to inconsistencies of one type or another. Inconsistencies may occur in adults’ informal speech, in children’s own ungrammatical utterances, and in the ungrammatical utterances of other learners such as playmates, siblings, and nonnative language learners. Inconsistencies also occur naturally in language. In English, many but not all verbs take the regular *-ed* ending for the past tense (“jumped” vs. “ran”), whereas in Spanish the gender of the determiner matches that of the noun in most but not all cases (“la tienda” vs. “el día”). In these instances, children must distinguish consistent from inconsistent or even idiosyncratic language forms. This is unavoidable in daily language. We propose that speech that serves as counterexamples to the therapy target (e.g., a new word, a regular morphosyntactic form) can interfere with learning in treatment contexts. In other words, when learning is implicit, even the typically developing learner has a limited ability to sort between the good input and the bad.

There is experimental evidence that typically developing learners can tolerate a low number of counterexamples during training and still generalize a rule. Gómez and Lakusta (2004) demonstrated that typically developing infants generalized grammatical patterns only when relatively low levels of counterexamples were present in the input (i.e., counterexamples comprised 17% of the total input). In contrast, no learning occurred when counterexamples comprised as little as 33% of the total input provided. This suggests that typically developing infants can track regularities in probabilistic input as long as the number of counterexamples is held to a relatively low level.

We have little information concerning how well children with developmental language disorders deal with exceptions to the dominant language patterns of their language, but there is reason to believe that counterexamples pose a problem for these learners. As Leonard, Camarata, Brown, and Camarata (2004) point out, typically developing children go through a period during which they mix up then resolve uses of forms such as the third-person singular *-s* (he plays vs. they play), *is* and *are* as copulas (Mollee *is* my friend vs. John and Jill *are* my friends) and auxiliaries (The boy *is* talking vs. The boys *are* talking), and the past *-ed* (producing grammatical *slept* vs. ungrammatical *sleeped*). Children with developmental language disorders take much longer than typically developing children to resolve this phase,

suggesting they have difficulty in separating the dominant pattern from the exceptions.

In treatment, children who receive input that includes both inflected and uninflected verb forms can show poor or no learning. Fey and Loeb (2002) did not produce gains when children were provided with recasts that did not provide direct models of the grammatical form being trained. Likewise, Fey and Loeb (2002) and Fey, Leonard, Bredin-Oja, and Deevy (2017) showed that children who heard input that included both third-person singular markings and the same verbs in unmarked form showed much poorer acquisition of this form compared to those who consistently heard the marked form of the verbs. There is one example in which counterexamples were intentionally included in a treatment context. Leonard et al. (2004) and Leonard, Camarata, Pawłowska, Brown, and Camarata (2006) intentionally employed counterexamples to convey the idea that certain grammatical forms cannot be omitted. As part of a larger treatment plan, a group of children heard counterexamples to the correct grammatical form. For example, children heard “Do you know where Bobby’s grandmother lives? She live\* on a farm. Whoops, I meant to say she lives on a farm!” (Leonard et al., 2004, p. 1370). Because these training episodes were part of a study with multiple forms of training episodes, we cannot know the unique effect of these counterexamples, making it difficult to know whether exposure to ungrammatical forms interfered with learning in this case.

Importantly, new work suggests that infant learners may be able to segregate instances of counterexamples by talker voice. Gonzales, Gerken, and Gómez (in review) exposed typically developing 12-month-olds to artificial language streams representing different dialects in which one dialect represented counterexamples to the other dialect. Input reflected a mixture of a “pure stream,” with sentences adhering to only one dialect. The other was a “mixed stream,” adhering to a 50:50 ratio for the two dialects. Infants generalized to novel sentences when the two streams were presented by different talkers, but not when the same talker presented both streams. This has direct implications for group treatment settings. If children with developmental language disorder are equally able to segregate input by talker voice, then we would have more assurance that their learning might be guided by consistently grammatical talkers (i.e., clinicians rather than other impaired children). Robust segregation of input by talker would lower the risk that the ratio of faulty production of multiple children in group therapy to correct clinician productions would undermine learning in groups. Whether this is true remains to be determined.

In summary, counterexamples to the treatment target should be held to a bare minimum during treatment. When learning is implicit, incorrect examples can erode the nascent internal representation of the correct form. Purposefully providing counterexamples as a contrast to correct forms can undermine learning if enough counterexamples are provided. Counterexamples can also come from other children present in group treatment sessions. We currently lack definitive data on whether this is a significant concern. If it

is, the occurrence of counterexamples in group treatment sessions could be reduced by grouping children who have different types of language errors, reducing the number of counterexamples to any given child’s treatment target that will occur because of their therapy partners’ language errors.

## **Input Principle II: Input Alone Can Affect Output**

The basic paradigm used in a statistical learning study focuses on the nature of the input provided to the learner. This runs counter to a common clinical assumption that expressive practice is required to change expressive performance. However, there have long been counterarguments to this assumption. When children generate utterances for the first time, they must rely on an internal representation of how words and morphemes fit together into grammatically correct patterns. In treatment, many children with developmental language disorder only learn the examples provided during a therapy session but often do not generalize this training outside of treatment. This indicates that they encoded what they heard, but did not change their internal representation on which generalization relies. It is this internal representation that is formed through the cognitive processes involved in statistical learning (Erickson & Thiessen, 2015). We propose not only that carefully crafted input shifts learners’ internal representations but that these new representations can change their expressive language as well.

Early evidence that input can change output comes from a 1976 treatment study for children with language disorders (Courtright & Courtright, 1976). In a modeling condition, clinicians presented models of the pronoun “they” to children who had not acquired this form. In an imitation condition, children also heard models of “they” but were also required to imitate each model for expressive practice. Imitation produced better results after the first session, but children failed to make further progress with additional sessions. In contrast, children who just heard models showed superior progress as sessions accumulated. After only three sessions, these children were using “they” in novel contexts, generalizing correctly just over 70% of the time.

Subsequent experimental evidence backs up the idea that the input children hear can alter their use of particular morphosyntactic structures. Leonard and Deevy (2017) exposed two groups of children with developmental language disorder to one or the other of two similar syntactic forms. The children who heard a “was + verb + ing structure” (comparable to “The cat was purring”) were more likely to produce this same structure than children who heard an alternate “verb + ing” syntactic form. Children who heard “verb + ing” form as part of a nonfinite verb structure (comparable to “We heard the cat purring”) produced this verb form more often than forms that contained the auxiliary verb.

The input principle is not limited to morphosyntactic learning. It also influences learning of the phonological



patterns that compose new lexical labels. Richtsmeier, Gerken, Goffman, and Hogan (2011) demonstrated this principle in a task that required typically developing children to learn and produce novel animal names. Ordinarily, typically developing children learn lexical labels best when the phonological composition of the word reflects patterns that appear frequently in their native language (i.e., high phonotactic frequency). Richtsmeier et al. (2011) asked whether learning words that reflected more difficult phonotactic patterns could be enhanced by increasing their frequency of occurrence within a training session. They demonstrated that learning low English phonotactic frequency words could be improved when typically developing children (a) heard new lexical labels frequently within the context of the study and (b) heard these frequently presented words spoken by multiple talkers. Plante, Bahl, Vance, and Gerken (2010) confirmed that high-frequency, multitalker input also benefited novel word productions for children with developmental language disorder. These types of studies provide evidence that expressive practice is not critical for implicit language learning. This makes statistical learning approaches particularly viable for low verbal children or children who need time to be comfortable in a new treatment setting before producing utterances on their own.

### **The Memory Principle: Learned Patterns Must Be Coded in Memory to Support Later Use**

Statistical learning is thought to require several cognitive processes, one of which is the ability to encode the patterns in the input that reflect new language forms in to memory (Thiessen, 2017). Unfortunately, there is ample evidence that children with developmental language disorders have verbal memory limitations. Several studies have suggested that children and adults with developmental language disorders do not encode information as well as their peers (Alt, 2011; Alt & Suddarth, 2012; McGregor, Gordon, Edem, Arbozo-Kelm, & Oleson, 2017). In addition, they appear to have deficits in long-term memory formation as they do not consolidate (or stabilize and retain) memories as well as their typically developing peers. This is true not only for verbal materials (Alt & Spaulding, 2011; Kuppuraj, Rao, & Bishop, 2016) but also when performing implicit statistical learning (Desmottes, Meulemans, & Maillart, 2016; Hedenius et al., 2011).

Poor encoding of recently learned material can be assisted clinically through processes of repeated retrieval. An enduring finding in the memory literature is that the mere act of retrieval strengthens newly encoded memories (Roediger & Karpicke, 2006). Repeated retrieval can be incorporated into training as a means of stabilizing the encoding of a memory through retrieval. Vlach and Sandhofer (2012) demonstrated this effect in a study of word learning with typically developing 3.5-year-old children. Children who were asked to produce a newly learned word label during training, “Can you say *koba*?” were more likely to retain that word a month later, whereas memory eroded significantly

for children who had not produced the word. This suggests a specific role for “correct” child productions in facilitating learning. Similarly, McGregor et al. (2017) showed that delayed retention of novel words could be improved for adults with and without developmental language disorder simply by testing their free recall between memorization trials.

Newly encoded memories are labile and prone to revision when reactivated (Gómez, 2006; Hupbach, Gómez, Hardt, & Nadel, 2007; Nader, 2003). Information encountered at reactivation that is consistent with prior learning can strengthen a prior memory. This happens via a process of reconsolidation of the reactivated memory. Information inconsistent with prior learning may weaken or update the reactivated memory via reconsolidation of the reactivated memory with the new information. In other words, bringing information encountered at an earlier time back into consciousness can help stabilize or revise that information in memory. Therefore, treatment components that cause a child to reactivate a treatment target provide an opportunity to reinforce any correct representations that are newly emerging. Perhaps even more important, reactivation and subsequent reconsolidation provide an opportunity for learners to alter their previous representation based on new input.

Certain established published treatments already incorporate procedures that facilitate reconsolidation. One example is the procedures involved in conversational recast treatment (Cleave, Becker, Curran, Owen Van Horne, & Fey, 2015). In conversational recasting, a child reactivates a memory for a language form heard in treatment by attempting to produce that particular form on their own. This reactivation makes their memory for the language form labile and open to change. The clinician immediately recasts the child’s utterance using the correct form. This correct form can act to override and update the child’s original memory. The child then consolidates the updated memory, and this procedure repeats within and across sessions, leading to improvement. Evidence consistent with improvement through reconsolidation comes from a morphosyntactic treatment study that incorporated an opportunity for reactivation and updating (Plante et al., 2018). Children heard high-density input that contained their target morpheme (i.e., bombardment) either before or after a period of enhanced conversational recasting. In that study, more individual children showed a positive treatment response when bombardment occurred at the end of treatment, compared to when the bombardment phase preceded recasting. The final period of bombardment offered an extra opportunity, beyond that offered by recasting alone, for children to reactivate their memory for their treatment target and to modify that internal representation based on the models provided through bombardment.

Incorporating multiple opportunities for reactivation and updating of target behaviors during treatment is likely to stabilize children’s internal representations of the treatment goal. Yet, despite the potential utility of incorporating repeated recall and reconsolidation opportunities, there is actually scant evidence for how these strategies can be deployed in treatment. Although almost any method used to periodically refresh memory for trained targets is likely

to promote a repeated recall effect, we have only a few basic principles for implementing this technique. We know, for example, that when new information is immediately followed by additional information, the subsequent information can interfere with memory encoding (Dudai, 2004). Also, frequent requests for repetitions are not well tolerated by children with language impairment (Haley, Camarata, & Nelson, 1994), suggesting that repeated and explicit production demands may cause children to disengage from treatment. How frequent opportunities for recall should occur for optimal effect in treatment is yet unknown. For typically developing children, uttering a label for a newly exposed word one time was sufficient to promote recognition of the word referent a month later (Vlach & Sandhofer, 2012). Although this suggests a role for expressive productions in treatment, some caution is warranted. Incorrect child productions may serve to further ingrain language errors in memory. This may be why treating verb morphology with less frequently produced (i.e., less ingrained) verb forms appears to be more effective than presenting morphemes attached to verbs children produce frequently (Owen Van Horne, Curran, Larson, & Fey, 2018).

## Concluding Remarks

The central thesis presented here is that harnessing principles that have emerged from the large body of statistical learning and memory research has the potential to facilitate treatment outcomes for children with developmental language disorders. Application of these principles does not necessarily require the creation of new treatment methods from whole cloth. Instead, thoughtful modifications to existing treatments to incorporate important learning principles can speed learning and improve generalization. We have laid out the evidence for this thesis, from basic experimental research and from treatment research, where it exists. We have also been careful to note that clinical translation of the basic research is still in its early days by noting the boundaries between what is known and information that is still needed.

Filling the current gaps in knowledge will require a specific approach to treatment research. Many published treatment studies are designed to address the question, “Does this particular treatment method work?” From an evidence-based treatment perspective, this is a useful research question. There are hundreds of these types of studies, ranging from single-subject case studies to randomized clinical trials. Many involve an amalgam of different procedures, even though there is often little or no information on the efficacy of the individual components. However, addressing the critical gaps in how learning principles can be translated to treatment will require a different approach. These studies must go beyond “Does it work?” to questions like “What makes it work?” and “Of these procedural options, which makes treatment work best?” Answers to these critical questions require studies that contrast two or more conditions in which individual treatment components are independently manipulated (Fey & Finestack, 2009).

These types of studies, although they exist, are still few in number.

It is worth restating that this discussion has purposefully sidestepped the issue of whether more naturalistic or more structured or even drill-like treatments are preferable as treatment methods. As applications of statistical learning principles move forward, these debates concerning optimal treatment contexts may prove to be secondary. It is true that a clinician is unlikely to get a toddler to cooperate with drill-like or even highly structured or scripted therapies. Likewise, adults seeking language treatment to address academic or vocational challenges are likely to prefer treatment contexts that mimic the type of work they need to improve rather than treatment solely in social conversational contexts. However, the evidence to date suggests that statistical learning principles can be applied fruitfully across a number of different educational or treatment contexts. For example, Alt et al. (2014) incorporated these principles into free-play. Plante and colleagues have used a method they refer to as enhanced conversational recast that uses a variety of age-appropriate activities but also imposes a structure that includes attentional cues and prompts for child productions that happen at rates much higher than would occur in natural conversations. Apfelbaum et al. (2013) and Powell et al. (2015) used worksheets to train spelling and math concepts. The variety of treatment contexts in which statistical learning principles have been applied suggests that decisions concerning the treatment context might be best addressed by considering the learning objective and the client’s age. After this, the specific procedures used within this context should conform to the principles outlined here that are known to enhance learning. This recommendation shifts the emphasis from the overt acts performed by the clinician and client during treatment to the nature of the input the clinician provides (see also Leonard & Deevy, 2017, for further discussion of this idea).

A final emphasis on the implicit nature of statistical learning is warranted. Explicit teaching methods do not engage implicit learning mechanisms. Indeed, explicit teaching can often inadvertently produce conditions that could actually impede learning. For example, explanations to children may require knowledge of vocabulary that is beyond those with a developmental language impairment. Asking children to think about language explicitly will require metalinguistic skills that may not be age appropriate, even for typically developing children, much less for those with disorders. Furthermore, explanations that center around a few examples will not permit learners to detect broader the underlying patterns that permit generalization to untrained examples. By intentionally avoiding explicit teaching in favor of implicit learning, clinicians can harness the cognitive resources that support rapid learning. Those discussed here include detection of sound sequences for words, detection of co-occurring events (e.g., objects and their labels), and inferring class membership (for morphosyntactic forms) and memory formation processes that help to retain these pieces of information. Although not covered in detail here, basic attention to the input is also critical. Even optimally provided input

is useless if the client is not attending to the input when it is provided (Meyers-Denman & Plante, 2016). Harnessing these processes through implicit statistical learning produces rapid learning that generalizes beyond the specific examples trained.

## Acknowledgments

Work involving translation of statistical learning principles to treatment by these authors is supported by NIDCD Grant R01DC015642 (E. Plante, principle investigator; R. Gómez, co-investigator). Portions of this review article were presented at the Callier Center, University of Texas, Dallas, in 2015 and at the Symposium for Research in Child Language Disorders in 2016.

## References

- Aguilar, J. M., & Plante, E. (2014). Learning of grammar-like visual sequences by adults with and without language-learning disabilities. *Journal of Speech, Language, and Hearing Research, 57*, 1394–1404.
- Aguilar, J. M., Plante, E., & Sandoval, M. (2018). Exemplar variability facilitates retention of word learning by children with specific language impairment. *Language, Speech, and Hearing Services in Schools, 49*, 72–84.
- Alt, M. (2011). Phonological working memory impairments in children with specific language impairment: Where does the problem lie? *Journal of Communication Disorders, 44*, 173–185.
- 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*, 487–498.
- Alt, M., Meyers, C. M., Oglivie, T., Nicholas, K., & Arizmendi, G. (2014). Cross-situational statistically based word learning intervention for late-talking toddlers. *Journal of Communication Disorders, 52*, 207–220.
- Alt, M., & Spaulding, T. (2011). The effect of time on word learning: An examination of decay of the memory trace and vocal rehearsal in children with and without specific language impairment. *Journal of Communication Disorders, 44*, 640–654.
- Alt, M., & Suddarth, R. (2012). Learning novel words: Detail and vulnerability of initial representations for children with specific language impairment and typically developing peers. *Journal of Communication Disorders, 45*, 84–97.
- Apfelbaum, K. S., Hazeltine, E., & McMurray, B. (2013). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology, 49*, 1348–1365.
- Arbel, Y., & Donchin, E. (2014). Error and feedback processing by children with specific language impairment—An ERP study. *Biological Psychology, 99*, 83–91.
- Braine, M. D. S. (1987). What is learned in acquiring words classes: A step toward acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 65–87). Hillsdale, NJ: Erlbaum.
- 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.
- Courtright, J. A., & Courtright, I. C. (1976). Imitative modeling as a theoretical base for instructing language-disordered children. *Journal of Speech and Hearing Research, 19*, 655–663.
- Desmottes, L., Meulemans, T., & Maillart, C. (2016). Later learning stages in procedural memory are impaired in children with specific language impairment. *Research in Developmental Disabilities, 48*, 53–68.
- Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology, 55*, 51–86.
- Eidsvåg, S. S., Austad, M., Plante, E., & Asbjørnsen, A. E. (2015). Input variability facilitates unguided subcategory learning in adults. *Journal of Speech, Language, and Hearing Research, 58*, 826–839.
- Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 134*, 406–419.
- Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66–108.
- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 321–335.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2006). *The MacArthur Bates Communicative Development Inventories*. Baltimore, MD: Brookes Publishing.
- Fey, M. E., & Finestack, L. (2009). Research and development in child language intervention: A five-phase model. In R. G. Schwartz (Ed.), *Handbook of child language disorders* (pp. 513–531). New York, NY: Psychology Press.
- Fey, M. E., Krulik, T. E., Loeb, D. F., & Proctor-Williams, K. (1999). Sentence recast use by parents of children with typical language and specific language impairment. *American Journal of Speech-Language Pathology, 8*, 273–286.
- 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.
- Fey, M. E., & Loeb, D. (2002). An evaluation of the facilitative effects of inverted yes–no questions on the acquisition of auxiliary verbs. *Journal of Speech, Language, and Hearing Research, 45*, 160–174.
- Fiser, J., & Aslin, R. N. (2002a). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*, 15822–15826.
- Fiser, J., & Aslin, R. N. (2002b). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458–467.
- Frank, M. C., Tenenbaum, J. B., & Gibson, E. (2013). Learning and long-term retention of large-scale artificial languages. *PLoS One, 8*, e52500.
- Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*, 218–245.
- Gerken, L. A. (2004). Nine-month-olds extract structural principles required for natural language. *Cognition, 93*, B89–B96.
- Gerken, L. A., Wilson, R., & Lewis, W. (2005). 17-month-olds can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249–268.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431–436.
- Gómez, R. L. (2006). Dynamically guided learning. In M. Johnson & Y. Munakata (Eds.), *Attention & performance XXI: Processes of change in brain and cognitive development* (pp. 87–110). New York, NY: Oxford University Press.



- Gómez, R. L., & Gerken, L.A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109–135.
- Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, 7, 567–580.
- Gonzales, K., Gerken, L. A., & Gómez, R. L. (in press). How who is talking matters as much as what they say to infant language learners. *Cognitive Psychology*.
- Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, 18, 254–260.
- Gray, S. (2004). Word learning by preschoolers with specific language impairment predictors and poor learners. *Journal of Speech, Language, and Hearing Research*, 47, 1117–1132.
- Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders*, 39, 158–170.
- Haley, A., Hulme, C., Bowyer-Crane, C., Snowling, M. J., & Fricke, S. (2017). Oral language skills intervention in preschool—A cautionary tale. *International Journal of Language and Communication Disorders*, 52, 71–79.
- Haley, K. L., Camarata, S. M., & Nelson, K. E. (1994). Social valence in children with specific language impairment during imitation-based and conversation-based language intervention. *Journal of Speech and Hearing Research*, 37, 378–388.
- Hall, J., Owen Van Horne, A., & Farmer, T. (2018). Distributional learning aids linguistic category formation in school-age children. *Journal of Child Language*, 45, 717–735.
- Hall, J., Owen Van Horne, A., McGregor, K. K., & Farmer, T. (2017). Distributional learning in college students with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 60, 3270–3283.
- Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veri'ssimo, J., Dye, C. D., . . . Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. *Research in Developmental Disabilities*, 32, 2362–2375.
- Hupbach, A., Gómez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory*, 14, 47–53.
- Kamhi, A. G. (1988). A reconceptualization of generalization and generalization problems. *Language, Speech, and Hearing Services in Schools*, 19(3), 304–313.
- Kamhi, A. G. (2014). Improving clinical practices for children with language and learning disorders. *Language, Speech, and Hearing Services in Schools*, 45(2), 92–103.
- Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83, B35–B42.
- Kittleson, M. M., Aguilar, J. M., Tokerud, G. L., Plante, E., & Asbjørnsen, A. (2010). Implicit language learning: Adults' ability to segment words in Norwegian. *Bilingualism: Language and Cognition*, 13, 513–523.
- Kuppuraj, S., Rao, P., & Bishop, D. V. M. (2016). Declarative capacity does not trade-off with procedural capacity in children with specific language impairment. *Autism & Developmental Language Impairments*, 1, 1–17.
- Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, 19(12), 1247–1252.
- Lany, J., Gómez, R. L., & Gerken, L. A. (2007). The role of prior experience in language acquisition. *Cognitive Science*, 31, 481–508.
- Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 47, 924–943.
- Law, J., Garrett, Z., & Nye, C. (2010). Speech and language therapy interventions for children with primary speech and language delay or disorder (Review). *The Cochrane Collaboration*. Hoboken, NJ: Wiley.
- 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, 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, 749–770.
- Leonard, L. B., & Deevy, P. (2017). The changing view of input in the treatment of children with grammatical deficits. *American Journal of Speech-Language Pathology*, 26(3), 1030–1041.
- Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77–80.
- McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of Speech, Language, and Hearing Research*, 60, 2891–2905.
- Meyers-Denman, C. N., & Plante, E. (2016). Dose schedule and enhanced conversational recast treatment for children with specific language impairment. *Language, Speech, and Hearing Services in Schools*, 47, 334–346.
- Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19, 498–550.
- Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with transformational cues to phrase structure. *Journal of Memory and Language*, 28, 360–374.
- Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, 20, 67–85.
- Nader, K. (2003). Memory traces unbound. *Trends in Neuroscience*, 26, 65–72.
- Opitz, B., Ferdinand, N. K., & Mecklinger, A. (2011). Timing matters: The impact of immediate and delayed feedback on artificial language learning. *Frontiers in Human Neuroscience*, 5, article 8.
- Owen Van Horne, A. J., Curran, M., Larson, C., & Fey, M. E. (2018). Effects of a complexity-based approach on generalization of past tense *-ed* and related morphemes. *Language, Speech, and Hearing Services in Schools*, 49(3S), 681–693. [https://doi.org/10.1044/2018\\_LSHSS-STLT1-17-0142](https://doi.org/10.1044/2018_LSHSS-STLT1-17-0142)
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80, 674–685.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, 113, 244–247.
- Perry, L. K., Samuelson, L., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability



- supports higher-order generalization and word learning. *Psychological Science*, 21, 1894–1902.
- Plante, E., Bahl, M., Vance, R., & Gerken, L. A.** (2010). Beyond phonotactic frequency: Presentation frequency effects word productions in specific language impairment. *Journal of Communication Disorders*, 44, 91–102.
- Plante, E., Gómez, R., & Gerken, L. A.** (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, 35, 453–462.
- Plante, E., Oglivie, 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, 1–16.
- Plante, E., Patterson, D., Gómez, R., Almryde, K. R., White, M. G., & Asbjørnsen, A. E.** (2015). The nature of the language input affects brain activation during learning from a natural language. *Journal of Neurolinguistics*, 36, 17–34.
- Plante, E., Patterson, D., Sandoval, M., Vance, C. J., & Asbjørnsen, A. E.** (2017). An fMRI study of implicit language learning in developmental language impairment. *NeuroImage Clinical*, 14, 277–285.
- Plante, E., Tucci, A., Nicholas, K., Arizmendi, G. D., & Vance, R.** (2018). Effective use of auditory bombardment as a therapy adjunct for children with developmental language disorders. *Language, Speech, and Hearing Services in Schools*, 49, 320–333.
- Plante, E., Vance, R., Moody, A., & Gerken, L. A.** (2013). What influences children’s conceptualization of language input? *Journal of Speech, Language, and Hearing Research*, 56, 1613–1624.
- Powell, S. R., Driver, M. K., & Julian, T. E.** (2015). The effects of tutoring with nonstandard equations for students with mathematics difficulty. *Journal of Learning Disabilities*, 48, 523–534.
- Reber, A. S.** (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 7, 317–327.
- Reber, A. S.** (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 81, 115–119.
- Reber, A. S.** (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118, 219–235.
- Reber, A. S., & Allen, R.** (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition*, 6, 189–221.
- Reeder, P. A., Newport, E. L., & Aslin, R. N.** (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, 66, 30–54.
- Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S.** (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*, 37, 106–122.
- Richardson, J., Harris, L., Plante, E., & Gerken, L.** (2006). Subcategory learning in normal and language learning-disabled adults: How much information do they need? *Journal of Speech, Language, and Hearing Research*, 49, 1257–1266.
- Richtsmeier, P. T., Gerken, L. A., Goffman, L., & Hogan, T.** (2011). Statistical frequency in perception affects children’s lexical production. *Cognition*, 111, 372–377.
- Roediger, H. L., III, & Karpicke, J. D.** (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255.
- Rom, A., & Leonard, L. B.** (1990). Interpreting deficits in grammatical morphology in specifically language-impaired children: Preliminary evidence from Hebrew. *Clinical Linguistics and Phonetics*, 4, 93–105.
- Saffran, J. R., Johnson, E., Aslin, R., & Newport, E.** (1999). Statistical learning of tonal structure by adults and infants. *Cognition*, 70, 27–52.
- Saffran, J. R., Newport, E. L., & Aslin, R. N.** (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621.
- Sandoval, M., Patterson, D., Dai, H., Vance, C. J., & Plante, E.** (2017). Neural correlates of morphology acquisition through a statistical learning paradigm. *Frontiers in Psychology*, 8, Article 1234.
- Thiessen, E. D.** (2017). What’s statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B*, 372, 20166056.
- Torkildsen, J. V. K., Dailey, N. S., Aguilar, J. M., Gómez, R. L., & Plante, E.** (2013). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar. *Journal of Speech, Language, and Hearing Research*, 56, 618–629.
- Twomey, K. E., Ranson, S. L., & Horst, J. S.** (2014). That’s more like it: Multiple exemplars facilitate word learning. *Infant and Child Development*, 23, 105–122.
- Valian, V., & Coulson, S.** (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, 27, 71–86.
- Vlach, H. A., & Sandhofer, C. M.** (2011). Developmental differences in children’s context-dependent word learning. *Journal of Experimental Child Psychology*, 108, 394–401.
- Vlach, H. A., & Sandhofer, C. M.** (2012). Fast mapping across time: Memory processes support children’s retention of learned words. *Frontiers in Psychology*, 3, Article 46.