Published in final edited form as:
J Commun Disord. 2014 ; 52: 207–220. doi:10.1016/j.jcomdis.2014.07.002.

Cross-situational statistically-based word learning intervention for late-talking toddlers

Mary Alt[1], **Christina Meyers**[2], **Trianna Oglivie**[3], **Katrina Nicholas**[4], and **Genesis Arizmendi**[5]
[1,2,3,4,5]University of Arizona

Abstract

Purpose—To explore the efficacy of a word learning intervention for late-talking toddlers that is based on principles of cross-situational statistical learning.

Methods—Four late-talking toddlers were individually provided with 7–10 weeks of bi-weekly word learning intervention that incorporated principles of cross-situational statistical learning. Treatment was input-based meaning that, aside from initial probes, children were not asked to produce any language during the sessions. Pre-intervention data included parent-reported measures of productive vocabulary and language samples. Data collected during intervention included production on probes, spontaneous production during treatment, and parent report of words used spontaneously at home. Data were analyzed for number of target words learned relative to control words, effect sizes, and pre-post treatment vocabulary measures.

Results—All children learned more target words than control words, and, on average, showed a large treatment effect size. Children made pre-post vocabulary gains, increasing their percentile scores on the MCDI, and demonstrated a rate of word learning that was faster than rates found in the literature.

Conclusions—Cross-situational statistically-based word learning intervention has the potential to improve vocabulary learning in late-talking toddlers. Limitations on interpretation are also discussed. Cross-situational statistically-based word learning intervention for late-talking toddlers

Introduction

Late-talking toddlers

Although there are slightly different definitions for this term, late-talking toddlers are children, typically between the ages of 24–35 months, who are late to develop language. Specifically, they have an expressive vocabulary delay. These children do not have a primary diagnosis to explain their poor language development (e.g., hearing loss, autism) and are often too young to be formally labeled with a primary language impairment. Approximately 13–19% of all children are late-talkers (e.g., Rice, Taylor, & Zubrick, 2008), although only 7% of kindergarten-aged children have a primary language impairment (Tomblin et al., 1997). Outcomes for late-talkers are debated, but there is evidence that even if these children are not diagnosed later with primary language impairment, language still

[1]Contact Author's Information: University of Arizona, 1131 E. 2nd St., Tucson, AZ 85721, 520-626-6180, Fax: 520-621-9901, malt@email.arizona.edu.

may be a relatively weak skill for them. Children who were late-talkers perform less well on language measures than non-late-talking peers (e.g., Rescorla, 2011). During toddlerhood, this relatively large group of children is characterized by an ineffective word learning system and typically has not received services to train word learning techniques, making them an appropriate population in which to study learning-theory-based word learning intervention.

Components of word learning

Word learning is a process. At a simplistic level, a child must map a label to a referent. However, beyond a fast-mapping stage (i.e., the initial mapping of label to referent), a child must learn to correctly extend the label to other valid exemplars of the referent's category. Problems may arise at the level of the label, in which the child shows difficulty correctly parsing the acoustic stream or formulating the correct motor behaviors to accurately represent and or produce the lexical label. A child must also continually build his or her semantic representation of the referent, learn increasing details about the referent, and tease apart the salient component of an abstract representation. This is what allows a child to know that it is okay to use the word "bunny" to label a live white rabbit and a tasty chocolate rabbit, but not a dog. As a child learns words, he or she must master all of these components.

Learning theory characterizes how learners acquire rules or patterns from their environment. Specifically, in statistical learning theory, researchers are interested in the type of learning that happens when people are not explicitly trying to learn. This type of learning is called unguided learning or implicit learning. The point of implicit learning is that people are able to recognize patterns in input and track those patterns to extract words (e.g., Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Traditionally, research on statistical learning has been highly decontextualized (e.g., people listening to strings of nonwords with no visual context). However, researchers who are interested in how statistical learning applies to the more complex real-life word learning environment have included the concept of cross-situational learning (e.g., Smith, & Yu, 2008). Cross-situational learning accounts for how people track statistical co-occurrence of label and referent across different contexts.

A model of early word learning developed by Yu and Ballard (2007) integrates the role of statistical learning of the language input that children hear. They propose that children track how often a word and a potential meaning co-occur across contexts. Children observe the contrasts of words and objects across situations, and in turn, derive the constant referent from many possibilities. To test their hypothesis, Yu and Ballard developed a computer simulation of word learning based on data from real child/adult interactions. The computer simulation demonstrated that learning was more efficient (i.e., the simulated learner was able to make more correct word-meaning associations) when the learner had access to the novel word in numerous contexts. These included both linguistic contexts (e.g., nice duck, that's a duck) and physical contexts (e.g., seeing a duck in contrast with other objects). Context was even more important than raw number of exposures.

This concept is mirrored in work with adults. Kachergis, Yu, and Shiffrin (2009) describe components that enhance statistical learning. Kachergis and colleagues determined that three factors were most responsible for success of cross-situational learning: word referent frequency, contextual diversity, and within-trial ambiguity. Word referent frequency related to the number of times that a word/referent pair was presented to a learner, with results showing that higher-frequency pairs were learned more easily than low-frequency pairs. Contextual diversity was characterized as the number of different pairs a particular word/ referent pair was presented with. For example, a salt shaker and the word ‘salt’ appearing alone would have no contextual diversity. ‘Salt’ (word/referent) presented only with a pepper shaker and the word ‘pepper’ would have low contextual diversity. ‘Salt’ paired with ‘pepper’, ‘salsa’, and ‘ketchup’ (and their respective referents) would have higher contextual diversity. This was not a single trial issue, but related to cumulative exposures. Thus, if ‘salt’ was only ever presented in a field of two (e.g., salt and X), as long as the X varied across trials, it would be considered high contextual diversity. Higher contextual diversity was associated with better word learning than low contextual diversity. Within-trial ambiguity was characterized as how many other possible referents a label was paired with (i.e., participants were exposed to three words/three objects versus four words/four objects) for a single trial. Lower within-trial ambiguity was associated with better word learning than higher within-trial ambiguity. Kachergis et al. noted that if two of these three factors were controlled, then the influence of the third would be mitigated. The idea is that, by providing high intensity and high variability input, the learner is given sufficient opportunities to track patterns in order to make the appropriate associations between a label and referent. In doing so, the learner comes to extract the most salient features of the label-referent pair, because those are the elements that are stable within the input. This idea is echoed in findings from Pan, Rowe, Singer, and Snow (2005) who used individual growth trajectory modeling to show that vocabulary production in toddlers from low-income homes was positively related to maternal lexical diversity and defined as how many different words mother used, but not to overall maternal talkativeness. This is an example of lexical diversity, but recall that contextual diversity is important too.

Cross situational learning in a clinical context

There are at least two important components of cross-situational statistical learning theory as it might be applied to clinical populations: time and effort. In terms of time, studies of implicit learning typically demonstrate participant learning after minutes of exposure in a single session (e.g., Saffran et al., 2008). Given the right type of input, implicit learning happens quickly. In contrast, in intervention, it seems that the more time, the better, both in terms of number of sessions and overall length of intervention. Currently, evidence for treatment of speech and language delays in children shows that longer interventions (i.e. more than 8 weeks) lead to better outcomes (Law, Garrett, & Nye, 2004) and greater numbers of individual treatment units are correlated with improved functional communication measures (Jacoby, Lee, Kummer, Levin, & Creaghead, 2002). Learning is not happening quickly in the context of therapy. There are many clinical populations for whom orientation to task and focus can be a challenge (e.g., young children, people with autism) and intervention based on implicit learning could be particularly beneficial to those populations.

Recently, we argued that people should incorporate principles of statistical learning theory into treatment (removed for review purposes). Here we present a treatment based on principles of cross-situational statistical learning theory to determine its effect on the word learning of late-talking toddlers.

Variability

The first principle of cross-situational learning theory we chose to incorporate was variability. We incorporated both high linguistic and contextual variability into the treatment design. To achieve high linguistic variability, clinicians provided nontelegraphic input, with the target word produced in a variety of linguistic contexts. We chose nontelegraphic input because, although disagreement exists on the use of telegraphic input with young children (e.g., Law, 2011; van Kleeck et al. 2010), recent empirical work does not support the use of telegraphic input (Bredin-Oja & Fey, 2013). Importantly, from a learning theory perspective, nontelegraphic language is essential to provide necessary linguistic variability. Therefore, grammatically complete utterances should not be abandoned just because they do not appear to be facilitating learning for children with language impairment or delay in their daily environment. Some of the foundations of word learning in typically-developing infants cannot be achieved if a child is not receiving multi-word input. For example, there is evidence that children take advantage of their knowledge of phonotactics and use this knowledge to identify word boundaries. They do this by segmenting sound sequences that are more likely to occur between words and chunking together sequences that are likely to occur within a word on the basis of these transitional probabilities (Saffran, Aslin, & Newport, 1996). Within the first year of life English learners are able to use statistical computations to help identify word boundaries and this ability is the foundation for lexical and syntactic learning (Jusczyk, Houston, & Newsome, 1999). In fact, 17-month-olds are better able to learn labels for new words if they have had practice with statistical word segmentation (Estes, Evans, Alibali & Saffran, 2007). It would be impossible to use distributional statistics if the input did not provide different linguistic units for the sounds to be distributed across. Single words do not provide enough distributional information for statistical learning to occur.

Recall that word learning also involves semantics and that contextual diversity is a critical factor in cross-situational learning. There is evidence that variability in learning referents allows for better mapping for toddlers, including generalization to novel object and improved vocabulary learning overall. Perry, Samuelson, Malloy, and Schiffer (2010) taught two groups of toddlers a set of words. One group was trained using a “tight” set of physical exemplars, which were very similar in appearance. The other group was trained using a variable set of exemplars, which all illustrated a common concept, but were different in appearance. Both groups of children learned the words they were taught. However, at one month post treatment, the children in the variable training set group were not only better at generalizing to new objects, but showed markedly faster rates of word learning than the children in the tight training set. Therefore, our treatment design included variable semantic contexts for presenting the input to the toddler.

Principles of treatment

The second principle we included was token frequency, which in our case, we refer to as intensity of input, specifically, the number of times the target is presented. However, the obvious issue with transferring learning principles demonstrated primarily in people without language impairments to people with language impairments is that, by definition, people with language impairments have difficulty learning language. Clearly, the amount of input that children with language impairment receive in their natural environment is inadequate for them to learn words at the same rate as unimpaired peers. However, evidence shows that impaired learners can learn when provided with adequate input (see Alt et al., 2012 for review). In terms of input, we focused on amount of input per session, although there are clearly many other ways to examine input (e.g., session duration). There are few specific references for exactly how much input impaired learners need to learn, but the principle we followed was that more would be better. From the available literature, a minimum number was selected to ensure that there were a sufficient number of presentations of each target word. However, given that the relationship between intensity of input and word learning is underexplored, we quantified input in order to examine the potential contribution of intensity to enhanced learning. We chose to demonstrate the efficacy of a cross-situational statistical learning-theory based intervention with late talking toddlers because they need rapid vocabulary development and, developmentally may not be ready to engage in highly structured treatment which requires significant output from participants.

The current study

We designed a word learning intervention for late-talking toddlers that incorporates high variability and high intensity –key principles of cross-situational statistical learning. Our question was whether or not delayed learners could learn words provided with an intervention that included these components in an enhanced dosage relative to a natural environment. Thus, we included multiple possibilities for learning through various mechanisms in our treatment design. This early efficacy study is an intentional ‘kitchen sink’ approach and is not intended to isolate specific factors that may individually contribute to the outcome.

Our hypothesis is that late-talking toddlers have not learned as many words as timely-talking peers because they have not yet learned successful and efficient ways of extracting the patterns of word/referent pairs from their environments. There is evidence that success parsing words from the speech stream at 18-months predicts later vocabulary growth, including which late talkers will have a vocabulary spurt (Fernald & Marchman, 2012). Our first prediction was that children would learn more target words presented during treatment than control words that had not been presented in treatment.

Our second prediction was that we would see growth in words other than the target words. This growth should occur at a faster rate than that seen in untreated late-talkers, as reported in the literature. Recall that cross-situational learning theory is intended to explain how people learn patterns or rules. The control words are a subset of all potential words in a child’s environment. The control words are a set of 5–10 out of hundreds of potential words to be learned. Logically, it would make sense that a child would learn words that are explicitly taught before she learns words that are not directly taught. However, if she is learning *how* to learn and overall vocabulary is expected to increase, the chance that those 5–10 chosen words will be the words the child learns is statically lower than the probability of the child learning any of the other hundreds of words in the unrestricted set. If people can extract patterns or rules, they can apply them in other situations and thus, have the ability to learn. If the cross-situational statistical learning aspect of the intervention was working, as opposed to some other component of the treatment (e.g., clinician effect), then we would also expect children to increase their use of novel words outside the realm of the treatment room. One complication is that children might also apply this pattern recognition ability to the control words, which were chosen for their salience. However, given that the control words were not targeted in treatment, the probability of them increasing at the same rate or to the same degree as the control words was unlikely.

If treatment and control words both improve equally, an obvious alternative interpretation could be made that the improvement was due to maturation. It was difficult to find an appropriate control other than vocabulary given that the majority of the participants (3 of 4) were not combining words or using morphology. If improvement was due to maturation, we would expect to see *equivalent* improvement in target *and* control words. We would also expect to see a rate of linguistic growth for novel words outside the treatment room in line with what is expected in the literature.

In sum, there are four possible outcomes for the experiment:

1. There is no improvement. We would expect to see little to no difference between the treatment and control targets. The rate of vocabulary growth for non-study words should be the same than the published rates for late talkers in the literature.

2. There is improvement that is due to maturation. We would expect to see equivalent growth in terms of the target and control words. The rate of vocabulary growth for non-study words should be the same as the published rates for either late talkers or vocabulary spurts in the literature.

3. There is improvement in the target words that reflects a treatment effect, but not from fully-realized principles of statistical learning. In this case, we would expect to see sizable effect sizes for treatment and improved usage of target words compared to control words. However, the rate of vocabulary growth for non-study words should be the same than the published rates for late talkers in the literature (e.g., Robertson & Ellis Weismer, 1999).

4. There is improvement in the target words that is related to principles of learning theory. Just as in scenario three, we would expect to see sizable effect sizes for treatment and improved usage of target words compared to control words. However, we would also expect to see improved rates of vocabulary growth for non-study words relative to the published rates for either late talkers or vocabulary spurts in the literature (e.g., Perry, Samuelson, Malloy & Schiffer, 2010)

We predict that we will see a pattern like that of scenario four, thus providing evidence that there are positive effects from incorporating principles of cross-situational statistical learning theory into vocabulary instruction.

Results from this study will add to the literature by demonstrating how principles of cross-situational statistical learning theory can be applied to treatment. Translational work examining how theoretical principles extend to clinical populations is relatively sparse. We will also test the effect these principles have on treatment when applied to late-talking toddlers' vocabulary development. Practically, it will provide clinicians with ideas about whether or not manipulating input in the way we propose has the potential to be effective.

Methods

Participants

Four participants (2 male and 2 female) ages 23 months to 29 months (mean age of 25.25 months) enrolled in and completed this study, which was approved by the Human Subjects Internal Review Board at the University of Arizona. All participants' parents referred their children to the study, or were referred by research labs or clinics at the University of Arizona. Inclusion was based on verification that the children were: 1) below expected performance in terms of productive word use, 2) could approximate words well enough to adequately judge future productions as word attempts, 3) had families who were willing to follow the treatment protocol. Two participants were not included in the study based on the third criterion.

Details on participants' ages and number of productive words pre-treatment are listed in Table 1. Although predicting later language difficulties from late-talking status is not an exact science, certain risk factors have been suggested (e.g., Olswang, Rodriguez, & Timler, 1998). These include parental concern (all participants) and family history of language/ learning disability (*n* = 2). In terms of family history, one participant had two relatives with diagnosed language learning disabilities. Another had an older brother who demonstrated a language delay that has since resolved. One of our participants was born prematurely, which can be a risk factor for later language difficulties (e.g., Luoma, Herrgard, Martikainen, & Ahonen, 1998). One participant had no identifiable risk factors. All families reported speaking English only at home.

The procedures to determine inclusion criteria included an interview with the families in which case history information was taken and parents described their child's current language skills. In all cases, parents were concerned about their child's language development. They filled out the MacArthur-Bates Communicative Developmental Inventories (MCDI) – 2nd Edition – Words and Sentences (Fenson et al., 2007). Heilmann, Ellis Weismer, Evans and Hollar (2005) reported that the most common method of identifying late talkers in the literature was the use of a cut-off of the 10th percentile for productive vocabulary on the MacArthur-Bates Communicative Development Inventories, Word and Sentences. All of our participants met this threshold. We corroborated the parents' reports of their child's limited language by collecting two language samples: one with the parent and child and another with the child and a researcher.

The procedures to determine the second inclusion criterion concerning intelligibility included an elicitation of imitative speech samples in order to rule out frank motor speech impairment, given that most of these children did not produce spontaneous productions. Considering the age of our participants, we expected them to have developmental phonological patterns. However, we needed to ensure that word approximations would be related to the target clearly enough to allow for inter-rater reliability. We used materials from the *Goldman-Fristoe Test of Articulation – 2nd Edition* (GFTA) (Goldman & Fristoe, 2000) to get an estimate of articulation skills. Most children did not complete the whole test, and given that we were not using this in a standardized manner, we did not use the norms associated with this measure. The procedures to determine the third inclusion criterion included parent understanding and follow-through of data tracking. Parents were given explanations of spontaneous productions and were asked to agree to consistently complete weekly logs of their child's spontaneous productions (formatted worksheets were provided by the clinicians) and a post-treatment MCDI.

Stimuli

The research team, in close collaboration with the families, chose between 5 and 10 words for each child to learn and a matching number of control words that would not be used during treatment. It is common practice in single-subject design to have different numbers of targets for individual participants based on participant needs (e.g., Kiran, Sandberg, & Abbott, 2009). All target and control words had to be words that were not in the child's productive vocabulary. The target and control words consisted of concepts that were not novel to the children (e.g., family members, familiar foods and drinks, common verbs/adjectives). Control words were chosen to be as closely matched to the target word as possible in terms of frequency, length, word type and salience to the family. Examples of all target and control words can be found in Appendix A. Given that these children had significant communicative needs (i.e., rarely used words), target words for each participant were selected individually with parent input in order to choose words that were functionally relevant to each family. We asked each family to choose a list of words they would most like their child to say and to identify situations in daily life in which they were experiencing the most communication breakdowns due to their child's limited use of words. Once a situation was identified, we discussed it and generated words that the family thought would be most helpful for their child to have in that situation. We explained to families that we needed to produce sets of words that were equivalent in terms of importance to them, difficulty (e.g., syntactic category, lexical and phonological familiarity), and salience. We chose both the target and control words from the lists generated by the families. For example, one family chose 'milk' and 'tea' as a pair. The child regularly drank milk, the parent regularly drank tea, and the child showed a real interest in the parent's drink. Thus for this family, this was an appropriate pair. Knowing the importance of factors like phonotactic probability (e.g., MacRoy-Higgins, Shafer, Schwartz, & Marton, 2014; Storkel & Lee, 2011), neighborhood density (e.g., Stokes, 2013; Storkel & Lee, 2011), and lexical familiarity (e.g., Stokes, 2010), we tried to keep these factors in consideration as we chose word pairs, but importance to the family and equivalent salience were the deciding factors in word selection. T-tests showed that all these factors were equivalent for all children with the exception of a phonotactic probability mismatch for WL03 (See Appendix A).

Procedures

Participants were seen in a clinic setting for 14 to 20 individual sessions which, for data analysis purposes, were grouped into weekly session pairs in order to provide stability over the course of a week. Parents and siblings were welcome to attend the sessions. We reminded parents at the start of treatment about the philosophy of the treatment (i.e., that it is input based), and that it was fine if their children were not speaking during the therapy sessions. We gave parents the option to end or extend their treatment after seven session pairs. We established seven as a minimum number of sessions and ten as a maximum, but allowed for some flexibility to accommodate families' individual schedules. For example, some families were not available for a full ten weeks (e.g., summer vacations).

In addition to the participant and parent (and sometimes a sibling), each session included a clinician and a data tracker. To increase variability, children had several clinicians and data trackers over the course of their treatment. Clinicians were graduate students (including one certified SLP) who were supervised by the first author, a licensed, certified speech language pathologist. Clinicians were trained on how to administer the treatment, and had the theoretical explanation for the treatment explained to them.

Each treatment session included an initial block of probes for all target and control words followed by a block of input of target words. At the beginning of the session, probe cards with photos of all target and control words were used in conjunction with spoken cloze phrases (e.g., "He drinks...") or questions (e.g., "What's this?") to elicit a spontaneous production. To ensure spontaneous production, clinicians never provided the child with a model (e.g., "Say X"). The same photo cards were used each time, but were presented in random order each session. The child's response (or lack thereof) was transcribed. If the production was unclear, the clinician and data tracker would confer to reach consensus on whether the attempt was a true approximation or not.

For the treatment block, which immediately followed the probe block, clinicians used a variety of activities that would highlight the target words and not the control words. In order to achieve variability, clinicians had to plan for at least five different ways to present each target word and include the word in different positions within an utterance. In order to meet intensity guidelines for this study, clinicians were instructed to provide input for at least three target words every session, with at least 64 exemplars of each word for every session. This number was created based on the work of Camarata, Nelson, and Camarata (1994) who found that young children with specific language impairment required an average of 64 models before they produced a grammatical target spontaneously. Although their work addressed grammar, it is one of the few in the literature to report specific intensity data. Other data published on rate suggests a recast rate of greater than .82 recasts per minute (Proctor-Williams, Fey, & Loeb, 2001) for grammatical structures and 192 presentations in 25 minutes for learning a transitive structure (Childers & Tomasello, 2001). With sessions scheduled to last a minimum of 20 minutes, and a maximum of 50 minutes, our guidelines would keep us within these rates.

In order to ensure treatment fidelity (i.e., meeting a minimum input of 64 models of a target word), the data tracker used a form with the target and control words and recorded a tally mark every time a target or control word was said. Every production was tallied regardless of who produced the word (e.g., parent, clinician, data tracker). The data tracker periodically alerted the clinician to the number of presentations of each target word to ensure the requisite number of exposures was achieved. Also, the data tracker would alert the clinician (or parent) if she began to use the control words too often, typically defined as more than five productions. Fidelity was calculated using the data from the data tracker to determine in how many sessions at least three target words were produced at least 64 times. The average fidelity rate was 97.67% with a range of 96.77% –100%.

Any production from the child during the treatment session was recorded by the data tracker on-line, during the session and noted as either spontaneous or imitative. Given that these were toddlers, words did not need to be produced with adult like form. They simply had to be approximations that were close enough to the target word to be recognizable as attempts at the target, and only the target, word. With the amount of input the children were receiving, it could be argued that any production of a target word would be delayed imitation. However, we counted productions as spontaneous when they were unsolicited and used in appropriate contexts. Most of our participants were not regularly imitating words (WL02 and WL07 did, but only when explicitly instructed to), thus most of their productions were spontaneous. Imitations were words produced when a child was prompted (e.g., some parents just could not help themselves), or uttered directly following an adult production without clear referent to the concept to which a word was linked.

We will differentiate our definition of spontaneous and imitative utterances with an example. Consider the following scenario. The clinician and child were pouring milk for different action figures and the clinician has been modeling the use of the word ‘milk’ in multiple forms. An example of a spontaneous utterance would be if the child picked up a cup with milk in it and said “milk”. An example of an imitative utterance might occur if the child said ‘milk’ right after the clinician, while looking directly at the clinician, not at the milk, and not using prosody to indicate a request for milk. In this situation, there is no indication that the word is being used in a meaningful way and thus would be labelled as imitation and not counted in the data set. Only spontaneous utterances counted towards data for production. The data tracker and clinician conferred during and after each session to ensure that the data tracker’s recordings of spontaneous productions were in alignment with the clinician’s perception of the child’s productions. Given that children had not produced these words before, any novel productions were typically perceived as important, and often commented upon. For example, a parent might say “Wow! That’s the first time she’s said milk.” or a clinician might recast the child’s utterance, “Yes, you said “milk”, He wants milk.” This made it easy for the data tracker to record the production, or for a discussion about whether or not the child’s production was actually a legitimate attempt, increasing the reliability of the data.

The final component of the procedure was the parent report. Parents were given a sheet to fill out to record their child’s usage of any words outside of the intervention sessions each week. Parents were reminded that we were only interested in spontaneous productions. At the end of each session pair, we collected the parent’s data log, and gave them a new sheet for the week.

At the end of 7 – 10 session pairs, we re-administered the language samples with parent and clinician and asked parents to fill out the MCDI- words and sentences.

Design and Data Analysis

This was an AB design, with replication across subjects. The independent variable was a list of words that were either targeted in treatment or not targeted (i.e., served as controls). Control words were semantically related to target words (e.g., If a target was ball, a control was block). The dependent variable was number of words produced by the child and was measured in several ways. Production of target and control words was measured using probes at the beginning of each treatment session, records of words produced during treatment, and parental report of word use at home. These data were used to calculate effect size, and to compare the number of words used by condition (i.e. target v. control). Further, overall pre-intervention and post-intervention production of words was measured using the MCDI, filled out by the same parent each time. This also allowed us to measure rate of words learned per week.

A consistent baseline for target/control words was established using two data points, and confirmed with a third. Word use was seen as dichotomous: either a child used a word, or they did not. Recall that these children were not typically producing spontaneous speech. Thus, the presence of a spontaneous utterance in a context in which the referent was clear was counted as a word production. Given that the children were expected to have developmental phonological patterns, production did not need to be in adult-like form. The first data point was the first MCDI; target and control words could not be reported as being used in the child's developmental history to date. The second data point was the initial evaluation in which two language samples were elicited from the child: one with a parent and one with a clinician. The confirmation was the probe, which was the first activity in the first treatment session. If a child used a target or control word in probe portion of the first treatment session, the baseline would be extended until we could establish a stable baseline for other words. We did not have to extend our baseline for any participants; at the start of treatment there was no evidence that they had ever produced either the treatment or control words.

In order to determine if our treatment worked, we started by using visual analysis. We would expect to see a faster and higher rise in production of target versus control words. We also calculated effect sizes. We could not use a traditional d statistic because, by definition, our baseline levels of target word production were zero, thus providing no variance for A1. Instead, we followed the effect size calculation used by Plante et al. (2014), which was based on Beeson and Robey (2006). To calculate this *d2* we subtracted the mean of the baseline (0) from the mean of the last three treatment sessions and divided by the variance of the last three treatment sessions. If there were productions but no variance (e.g., 1, 1, 1), we used the minimum standard deviation possible, given a whole number (i.e., .57735). We examined the data in several ways including looking at each type of data point individually (probe, in treatment production, at home production), as well as in combination. That allowed us to see how representative the home productions were to what was directly observed by clinicians and data trackers during treatment session.

To measure word learning outside the treatment room, we used pre- post- MCDI scores to calculate the number of words learned. We calculated the number of words learned per week by dividing that number by the child's actual number of weeks in the study. Finally, we checked to see if the child's percentile on the MCDI had changed compared to their initial percentile.

Last, because we planned to keep track of input, we planned to use descriptive data to measure how many productions a child would be exposed to in treatment before he or she produced a word spontaneously.

Results

Individual and group effect sizes

In order to determine if there were changes in the children's performance on target and control words, we began with visual inspection of the data (see Figure 1.). All four participants ended treatment producing more target words than control words, and with the exception of one data point for one child, consistently produced more target than control words throughout treatment. Three participants produced target words earlier than control words (WL01 week 1 v 7; WL03 week 1 v. 4; WL07 week 1 v. 5).

Additionally, we calculate *d2* effect sizes. Given that there was a zero baseline for all these words (i.e. children had never produced these words before), a single non-imitative report of production of the word counted as use of the word. Figure 1 shows target and control words used by each participant, along with corresponding *d2* scores for the target words. This figure includes productions that occurred in any of our data collection options: probe, in-treatment production, parent report of production at home. We calculated the *d2* statistic for target words for all productions.

There might be concerns about the level of control involved in parent reporting. Therefore, we looked at effect sizes for the probe alone, in-treatment production only, and a combination of the directly observed data points (probe plus in-treatment production). Instances in which fewer than four data points are reported are a result of limited variability. (See Table 2.)

Target words learned compared to control words

A second measure of the specificity of the treatment effect was to determine how many of the targeted words participants would use compared to control words. Evidence of use could come from probes, in-treatment usage, or usage reported at home. Table 3. shows the details of words learned in each condition. The average percentage of target words produced was 90.75% (SD = 10.75%) for all conditions and 81.5% (SD = 13.98%) for directly observed conditions (i.e., excluding parent report data). In contrast, the average percentage of control words produced was 38.25% (SD = 3.5%) in both conditions.

Evidence of generalization

We expected the treatment would allow children to learn how to learn words. Thus, we expected other types of gains besides just improvement on the target words. One metric of growth was pre-post measures of productive vocabulary on the MacArthur-Bates Communicative Developmental Inventories – 2nd Edition (Fenson et al., 2007). Table 1 shows these results. The average number of words learned per week was 21.60, with a range of 10.27 - 37.66. To calculate this we looked at the entire duration of the treatment (e.g., including holidays when we did not have sessions) instead of just the actual number of sessions.

Input Intensity

Given that intensity was a component of our treatment design, we wanted to quantify how much input children received. First, we examined how many presentations of a word children had before they produced that word. We started by examining data for target words children actually produced, of which there were 24. We looked to see in which weekly session pair a child first produced a word. We then added all the inputs for that word up to and including the weekly session pair in which the word was produced. The average amount of input per word was 704 exemplars, with a standard deviation of 542. For the two words that were never produced, the amount of input provided was 976 and 1138, thus it unlikely that these words were not produced due to lack of input.

Next, we calculated how many examples of target words were produced in each minute of a session, regardless of whether or not a child produced a word. The average rate of input was 9.66 exemplars per minute, with a standard deviation of 2.18. Individual ranges for both calculations are found in Table 3. Figure 3 shows the average rate of input for each participant across session pairs.

Discussion

Interpretation of findings

Theoretically, we predicted that by capitalizing on principles of cross-situational statistical learning, we would help children learn to learn words by extracting patterns of word/referent pairs from the treatment environment. Observationally, children learned more target than control words, suggesting that the treatment was effective and somewhat more effective for treated versus untreated words. Additionally, the generalization of word-learning strategies as evidenced by overall increase in vocabulary for all four participants suggests that they not only learned new words, but learned how to learn untaught words outside of the treatment room.

Comparison to the literature

There are few effect sizes for vocabulary intervention reported in the literature to give a context for these findings. One can be interpolated from data reported in Ellis Weismer et al. (1993). Like the current study, linguistic input was varied to allow for words to appear in different positions and to have different contexts for introduction of the words. We calculated *d2* scores for their three participants using the responses for the individual treatments production probes for the modeling treatment for each participant, as this was the treatment closest to our intervention. The average for the three participants was *d2* = 1.03. Kouri (2005) reports on two different treatment approaches (modeling and elicitation) for children with language and developmental delays. The average d scores in that study were less than 1.0, whether these authors examined a strict definition of acquisition (with criteria for phonetics, context, and production across consecutive sessions) or included looser definitions. In this context, our average effect size for directly observed productions (*d2* = 3.79) can be considered large.

One possible interpretation for our participants' learning is that it was a result of maturation, and the inevitable increase in the rate of vocabulary growth that children have at this age (e.g., Ganger & Brent, 2004; Mitchell & McMurray, 2008). This growth typically occurs during the last half of a toddler's second year. Although there is debate in the literature about whether or not this increase is a true growth spurt or explosion, and, if so, how it should be measured (e.g., Ganger & Brent, 2004), there is no question that the rate of growth increases. Ganger and Brent (2004) report that earlier studies defined a spurt as being equivalent to the acquisition of roughly five words over a one week period. Goldfield and Reznick (1990) used roughly the same definition, and found that their participants (aged 15 – 22 months) who were defined as having a spurt learned roughly eight new words per week. Mitchell and McMurray (2008) interpret Dale and Fenson's (1996) lexical development norms to average between 7–8 words per week for 17 –21 month olds, compared to 2.7 words per *month* for 11–15 month olds.

In terms of the rate of vocabulary development for late-talkers, Rescorla, Mirak, and Singh (2000) used parent report to examine vocabulary growth rates for 28 late-talking toddlers between the ages of 24 and 36 months. Presumably, these children did not receive language treatment during this longitudinal study. Rescorla et al. found that children did not develop in a uniform manner. Eleven of the children were faster word learners than the rest (*n* = 17), that is, they had learned 150 words by age 30 months (a 24 month target). Over the course of age 2;0 to 3;0, the average rate of words learned per week was 5.58 for the fast learners, and 3.58 for the slower learners. If examined in six month intervals to adhere more closely to the Mitchell and McMurray (2008) figures, the words per week rates across the year were 6.74 and 4.37 for the faster learners and 0.66 and 6.55 words per week for the slower learners. Fernald and Marchman (2012) also present data that followed a mixed group of late-talkers (*n* = 36) from 18 months to 30 months, including some children who later showed evidence of more rapid vocabulary growth. Their cut-off of a MCDI percentile of <20 is slightly higher than the 10^{th} percentile usually used in the late-talking literature. The average rate of words per week for their group of late talkers was 7.47 words per week. (Rates were not broken out for the faster and slower learners). These rates are just slightly slower than the rates proposed for the vocabulary increase for younger timely-talkers and make it unlikely that late-talker maturation would involve a 'super-spurt' to help them catch up to timely-talkers.

The late-talking toddlers in our study showed vocabulary growth gains averaging 21.60 new words per week, with all four children learning more than ten new words per week. This is more than three times the average of the untreated late-talking toddlers in Rescorla et al. (2000), more than twice the average reported in Fernald and Marchman (2012) and more than twice the average for younger timely-talkers (Mitchell & McMurray, 2008).

We can also compare our participants' rate of growth to the growth of late talkers in other treatment studies. Robertson and Ellis Weismer (1999) enrolled 21 late-talking toddlers in a comprehensive 12-week treatment program that 'emphasized vocabulary development and use of early 2- or 3-word combinations within a social context" (p. 1240). Rate of input for targets was not specified. One outcome measure of this comprehensive language treatment study was expressive vocabulary on the MCDI- Words and Sentences. The toddlers in their control group (n = 10) gained an average of 10 words over the 12 weeks (0.83 wds/wk), and the toddlers in the experimental group (n = 11) gained an average of 37 words (3.08 wds/wk). Compared to the rates presented above, our late-talking toddlers seem to be exhibiting accelerated rates of lexical growth. They are above the average of 10.71 words/week learned by the late-talking toddlers (n = 12) in a Hanen focused treatment approach (11 weeks + 3 weeks for post-testing) and the toddlers in that study's control group (n = 13) (2.93 wds/wk, given a 4 month control phase) (Girolametto et al., 1996). This makes it unlikely that the growth our toddlers experienced is typical, and simply a result of maturation.

Contributions to growth

Our findings suggest that using a model of word learning intervention based on principles of cross-situational statistical learning theory has the potential to positively influence the vocabulary growth of late-talking toddlers. This was a first attempt to examine this hypothesis in practice, and thus there is much work to do to fully understand the implications of these findings. For example, we simultaneously provided a high level of variability of input (in language, activities, and clinicians), and intensity. Therefore, it is difficult to know which one of these measures, or the interactions between them, may be driving the gains we are seeing. It is also clear that not all children responded in precisely the same way, so individual differences may be contributing to different outcomes.

Our data do not allow us to speculate on the precise role of high variability. Rates of input are often not reported in the word learning literature, suggesting they are not often considered. However, our rate of input is far higher than rates of input that are reported in the literature. We set a minimum threshold of at least 64 examples of at least three targets for each session based on what was available in the literature. We easily surpassed that minimum, but were not actively monitoring just how much above the minimum each child's input was. Intensity may well be an important component of growth (e.g., Proctor-Williams et al, 2001), and the amount of input we provided children was substantially larger than what is typically provided in the literature. For example, other treatments for late talkers have included roughly five (Girolametto, Pearce, & Weitzman, 1996) or eight to ten models of each target word for each session (Ellis Weismer, Murray-Branch, & Miller, 1993). However, this relation between intensity and outcomes needs to be directly tested. Our data do not allow us to disentangle the precise mechanisms of what contributed to the word learning growth we saw in our participants.

Limitations

As mentioned earlier, this 'kitchen sink' approach to intervention does not allow us to specify which manipulations or interactions between manipulations may be responsible for the improvement we saw in our participants. However, this was a first attempt to examine the feasibility of this approach. We replicated the effect across four participants, but were not able to tightly control the timing of recruitment and intervention, so that we did not have a true multiple baseline across subjects design. We feel this is mitigated by the fact that we chose words that were not used in any situations by the participants. In other words, it was clear that all these children were delayed in their language development, parent report indicated ongoing concerns, and the participants did not appear to be on the cusp of maturation, based on the ease of finding and maintaining a stable baseline of zero productions for all participants.

This study was specifically designed for late-talking toddlers. It happens that all of our toddlers presented with age-appropriate receptive language skills (according to parent report), and although not assessed formally, there were no concerns about general non-verbal cognitive skills. However, we cannot assume that the pattern of learning we observed across our four toddlers will generalize to all other late-talkers, or to children with different profiles. Further work is needed to a) isolate the factors that influenced vocabulary learning in this study and b) determine if different types of intervention design are needed for children who have vocabulary deficits but do not fit a classic late-talker profile.

Summary

Using an input-based word-learning intervention that includes high variability of semantic and linguistic input with high intensity appears to jump-start the vocabulary growth of late-talking toddlers, when administered twice a week for 7 –10 weeks. The intensity of targets presented per minute used in this program was much greater than what has previously been reported in the literature. More work is needed to parse the specific mechanisms driving the improvement and to determine how this approach would apply to different populations.

Acknowledgments

The work presented in this paper was partially funded by support from the Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Research Training Grants (T32DC009398) from the National Institute for Deafness and Other Communication Disorders to the second and third authors and by support from the Initiative for Maximizing Student Development Program (IMSD) from the National Institutes of Health (2R25GM062584-12) for the fifth author. We would also like to acknowledge all the participants and their families who took part in the study, the members of the L4 Lab for their help with data collection, and the University of Arizona clinical faculty and staff for partnering with us on personnel and space.

References

Alt M, Meyers C, Ancharski A. Using principles of learning to inform language therapy design for children with specific language impairment. International Journal of Language & Communication Disorders. 2012; 47:487–498. [PubMed: 22938060]

Beeson PM, Robey RR. Evaluating single-subject treatment research: Lessons learned from the aphasia literature. Neuropsychology Review. 2006; 16:161–169. [PubMed: 17151940]

Bredin-Oja SL, Fey ME. Children's responses to telegraphic and grammatically complete prompts to imitate. American Journal of Speech-Language Pathology. 2014; 23:15–26. [PubMed: 24018697]

Camarata SM, Nelson KE, Camarata MN. Comparison of conversational- recasting and imitative procedures for training grammatical structures in children with specific language impairment. Journal of Speech, Language and Hearing Research. 1994; 37:1414–1423.

Childers JB, Tomasello M. The role of pronouns in young children's acquisition of the English transitive construction. Developmental Psychology. 2001; 37:739–748. [PubMed: 11699749]

Dale PS, Fenson L. Lexical development norms for young children. Behavior Research Methods, Instruments, & Computers. 1996; 28:125–127.

Ellis Weismer S, Murray-Branch J, Miller JF. Comparison of two methods for promoting productive vocabulary in late talkers. Journal of Speech, Language and Hearing Research. 1993; 36:1037–1050.

Estes KG, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science. 2007; 18:254–260. [PubMed: 17444923]

Fenson, L.; Marchman, VA.; Thal, DJ.; Dale, PS.; Reznick, JS.; Bates, E. MacArthur-Bates Communicative Development Inventories. 2. Paul H. Brookes Publishing Co; Baltimore, MD: 2007.

Fernald A, Marchman VA. Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late – talking toddlers. Child Development. 2012; 83:203–222. [PubMed: 22172209]

Ganger J, Brent MR. Reexamining the vocabulary spurt. Developmental Psychology. 2004; 40:621–632. [PubMed: 15238048]

Girolametto L, Pearce PS, Weitzman E. Interactive focused stimulation for toddlers with expressive vocabulary delays. Journal of Speech, Language and Hearing Research. 1996; 39:1274–1283.

Goldfield BA, Reznick JS. Early lexical acquisition: Rate, content, and the vocabulary spurt. Journal of child language. 1990; 17:171–183. [PubMed: 2312640]

Goldman, R.; Fristoe, M. Goldman-Fristoe Test of Articulation. 2. American Guidance Service, Inc; Circle Pines, MN: 2000.

Heilmann J, Weismer SE, Evans J, Hollar C. Utility of the MacArthur-Bates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. American Journal of Speech-Language Pathology. 2005; 14:40–51. [PubMed: 15966111]

Jacoby GP, Lee L, Kummer AW, Levin L, Creaghead NA. The number of individual treatment units necessary to facilitate functional communication improvements in the speech and language of young children. American Journal of Speech-Language Pathology. 2002; 11:370.

Jusczyk PW, Luce PA. Infants' sensitivity to phonotactic patterns in the native language. Journal of Memory and Language. 1994; 33:630–645.

Kachergis, G.; Yu, C.; Shiffrin, RM. In: Taatgen, N.; van Rijn, H.; Nerbonne, J.; Schomaker, L., editors. Frequency and contextual diversity effects in cross-situational word learning; Proceedings of 31st Annual Meeting of the Cognitive Science Society; Austin, TX: Cognitive Science Society; 2009. p. 755-760.

Kiran S, Sanberg C, abbot K. Tretment for lexical retrieval using abstract and concrete words in person with aphasia: Effect of complexity. Aphasiology. 2009; 23:835–853. [PubMed: 19816590]

Kouri TA. Lexical training through modeling and elicitation procedures with late talkers who have specific language impairment and developmental delays. Journal of Speech, Language and Hearing Research. 2005; 48:157–171.

Law J. Grammatical input is best for most children in the early stages of language development but telegrammatic input has its place for some. Evidence-Based Communication Assessment and Intervention. 2011; 5:11–14.

Law J, Garrett Z, Nye C. The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. Journal of Speech, Language and Hearing Research. 2004; 47:924–943.

Luoma L, Herrgard E, Martikainen A, Ahonnen T. Speech and language development of children born at 32 weeks' gestation: a 5 – year prospective follow – up study. Developmental Medicine & Child Neurology. 1998; 40:380–387. [PubMed: 9652779]

MacRoy-Higgins M, Shafer VL, Schwartz RG, Marton K. The influence of phonotactic probability on word recognition in toddlers. Child Language Teaching and Therapy. 2014; 30:117–130.

Mitchell, CC.; McMurray, B. A stochastic model for the vocabulary explosion. Proceedings of the 30th Annual Conference of the Cognitive Science Society; 2008. p. 1919-1926.

Olswang LB, Rodriguez B, Timler G. Recommending intervention for toddlers with specific language learning difficulties: We may not have all the answers, but we know a lot. American Journal of Speech-Language Pathology. 1998; 7:23–32.

Pan BA, Rowe ML, Singer JD, Snow CE. Maternal correlates of growth in toddler vocabulary production in low – income families. Child Development. 2005; 76:763–782. [PubMed: 16026495]

Perry LK, Samuelson LK, Malloy LM, Schiffer RN. Learn locally, think globally exemplar variability supports higher-order generalization and word learning. Psychological science. 2010; 21:1894–1902. [PubMed: 21106892]

Plante E, Oglivie T, Vance R, Aguilar JM, Dailey NS, Meyers C, Lieser AM, Burton R. Variability in the language input to children enhances learning in a treatment context. American Journal of Speech Language Pathology. 2014 (released online April, 3, 2014). 10.1044/2014_AJSLP-13-0038

Proctor-Williams K, Fey ME, Loeb DF. Parental recasts and production of copulas and articles by children with specific language impairment and typical language. American Journal of Speech-Language Pathology. 2001; 10:155–168.

Rescorla L. Late talkers: do good predictors of outcome exist? Developmental Disabilties Research Reviews. 2011; 17:141–150.

Rescorla L, Mirak J, Singh L. Vocabulary growth in late talkers: lexical development from 2;0 to 3;0. Journal of Child Language. 2000; 27:293–311. [PubMed: 10967889]

Rice ML, Taylor CL, Zubrick SR. Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. Journal of Speech, Language and Hearing Research. 2008; 51:394.

Robertson SB, Weismer SE. Effects of treatment on linguistic and social skills in toddlers with delayed language development. Journal of Speech, Language and Hearing Research. 1999; 42:1234–1248.

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science. 1996; 274:1926–1928. [PubMed: 8943209]

Saffran J, Hauser M, Seibel R, Kapfhamer J, Tsao F, Cushman F. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. Cognition. 2008; 107:479–500. [PubMed: 18082676]

Saffran JR, Newport EL, Aslin RN, Tunick RA, Barrueco S. Incidental language learning: Listening (and learning) out of the corner of your ear. Psychological Science. 1997; 8:101–105.

Smith L, Yu C. Infants rapidly learn word-referent mappings via cross-situational statistics. Cognition. 2008; 106:1558–1568. [PubMed: 17692305]

Stokes SF. Neighborhood density and word frequency predict vocabulary size in toddlers. Journal of Speech, Language, and Hearing Research. 2010; 53:670–683.

Stokes SF. The impact of phonological neighborhood density on typical and atypical emerging lexicons*. Journal of Child Language. 2013:1–24.

Storkel HL, Lee SY. The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. Language and Cognitive Processes. 2011; 26:191–211. [PubMed: 21643455]

Tomblin JB, Records NL, Buckwalter P, Zhang E, Smith E, O'Brien M. Prevalence of Specific Language Impairment in kindergarten children. Journal of Speech, Language, and Hearing Research. 1997; 40:1245–1260.

van Kleeck A, Schwarz AL, Fey M, Kaiser A, Miller J, Weitzman E. Should we use telegraphic or grammatical input in the early stages of language development with children who have language impairments? A meta-analysis of the research and expert opinion. American Journal of Speech-Language Pathology. 2010; 19:3–21. [PubMed: 19644126]

Yu C, Ballard DH. A unified model of early word learning: integrating statistical and social cues. Neurocomputing. 2007; 70:2149–2165.

Appendix A. Target and Control Words per participant

Subject # WL01	Target Words	Control Words
	BALL	BLOCK
	DOG	COW
	MOMMY	PAPA
	SHOE	BIB
	MOON	ROCK
	UP	ON
	WATER	SODA
	NANA	UNCLE TOM
	CAR	VAN
	GO	OFF
Average Biphones	.0046	.0034
Mean Log Frequency of neighbors	2.34	2.62
Average # of Neighbors	11.2 (2–22)	8.5 (1–16)
Average Log_Freq_HAL	10.91	10.09

Subject # WL02	Target Words	Control Words
	GIVE	TAKE
	WANT	NEED
	HUNGRY	THIRSTY
	DRY	FULL
	NOT	SOME
Average Biphones	.0051	.0053
Mean Log Frequency of neighbors	2.96	2.57
Average # of Neighbors	6.6 (0–18)	11.6 (0–17)
Average Log_Freq_HAL	11.77	11.65

Subject # WL03	Target	Control
	Words	Words
	GIVE	TAKE
	WANT	ASK
	HELP	SHOW
	MILK	TEA
	MINE	HERS
Average Biphones*	.0057	.0021
Mean Log Frequency of neighbors	2.50	2.71
Average # of Neighbors	7 (3–19)	14.6 (2 –27)
Average Log_Freq_HAL	11.84	10.77

* significant between group difference $p < .01$

Subject # WL07	Target Words	Control Words
	MATTHEW (Family name)	REGAN(Family name)
	HELP	FALL
	ANDREW(Family name)	TARYN(Family name)
	EAT	DRINK
	TIRED	WAKE
	LOOK	SAY
Average Biphones	.0029	.0045
Mean Log Frequency of neighbors	2.385	2.624
Average # of Neighbors	7.16 (0–20)	9.33 (0–24)
Average Log_Freq_HAL	10.97	10.70

Note: Calculations were made using the following tools:

For phonotactic probability and neighborhood density based on a child corpora: http://wwwbncdnetkuedu/cgi-bin/DEEC/post_cccvi

Storkel, H. L. & Hoover, J. R. (2010). An on-line calculator to compute phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behavior Research Methods*, *42(2), 497–506.*

For lexical frequency information and additional neighborhood density information: http://elexicon.wustl.edu/query14/query14domore.asp

Balota, D.A et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.

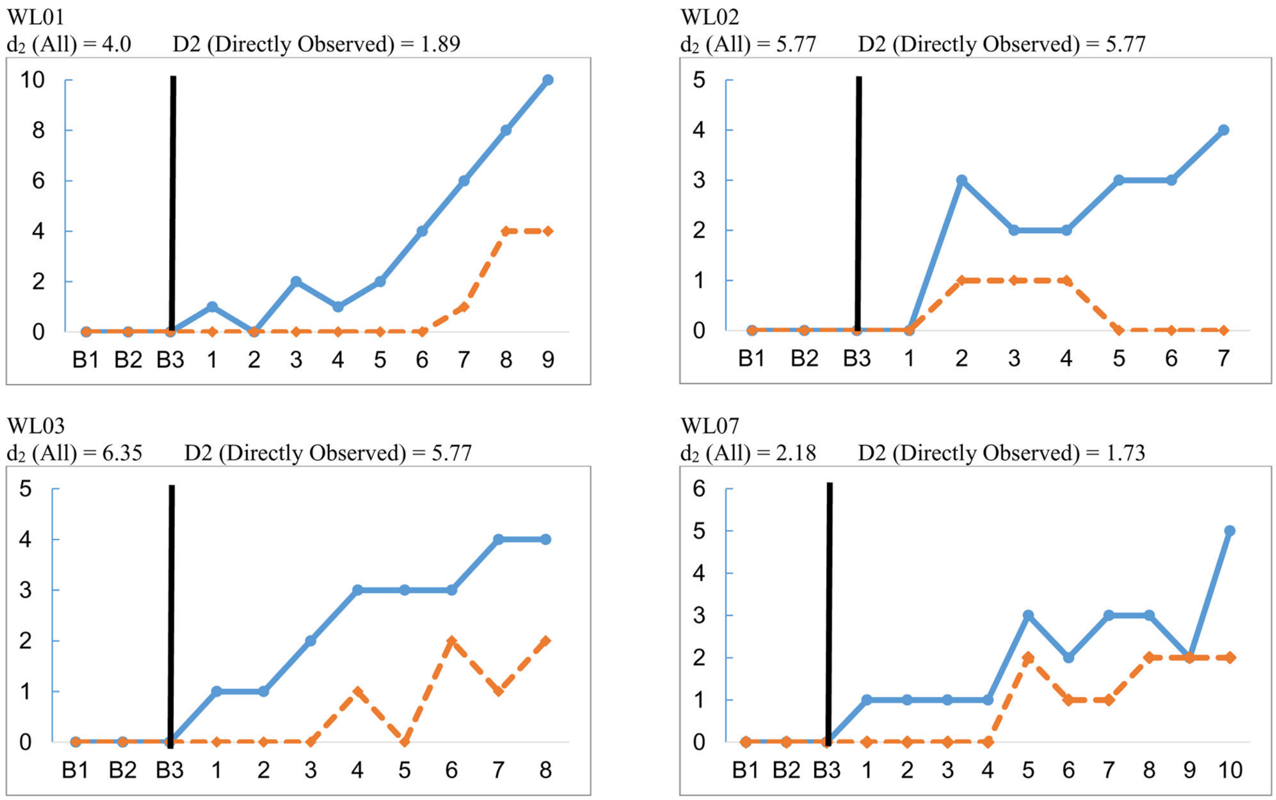

Figure 1.

Participant learning on target and control words across session pairs.

Notes: The x-axis represents baseline sessions and, after the black line, session pairs. The y-axis represents number of words to be learned. The solid lines represent target words produced, and the dashed lines represent control words produced. These lines represent productions from probe, in treatment, or home reporting. All = all possible productions (probe, in treatment, and home). Directly Observed = productions directly observed by clinicians (probes and in treatment).

Table 1

Information about participants and number of words produced on the MCDI pre- and post- treatment.

Participant		Age in In Months	Words Produced	Percentile	Combining Words?	Risk Factors
WL01	PRE	25	0	<5th	NO	Family History
	POST	27	339	25th	YES	
WL02*	PRE	29	209 (25)	10th–15th (<5 th)	YES	Premature Birth
	POST	33	545	55th	YES	
WL03	PRE	24	38	<5th	NO	None
	POST	27	263	15th–20th	YES	
WL07	PRE	23	48	5th– 10th	NO	Family History
	POST	25	161	20th –25th	YES	

MCDI= MacArthur-Bates Communicative Development Inventories- 2nd Edition, Words and Sentences

*The parent reported that the child produced 209 words pre-treatment, but clarified that only 25 were non-imitative. At post-test, this child was 33 months, and the norms for the MCDI only go to age 30 months. Thus, the post-test percentile is based on the 30 month norms, and is likely lower for a 33 month old.

Table 2

d_2 effect sizes for control and target words, averaged across participants

	N	Target	N	Control
All Conditions	4	4.57	3	2.69
Probe Only	3	1.31	2	2.02
In Tx Only	4	3.80	3	0.76
Directly Observed	4	3.79	3	1.91

[*] Note: All Conditions includes parental report, probe, and in treatment productions. In Tx Only is in treatment only productions. Directly observed includes Probe productions and in treatment productions combined.

Table 3

Number of words learned in the target and control conditions by participants

Participant	# of words to learn*	# Target Words produced (Total, Probe, In-Tx, Home)	# Control Words Produced (Total, Probe, In-Tx, Home)
WL01	10	10, 8, 8, 9	4, 4, 3, 2
WL02	5	4, 3, 4, 4	2, 2, 1, 0
WL03	5	5, 1, 5, 4	2, 0, 2, 2
WL07	6	5, 4, 3, 4	2, 2, 1, 1

*The number of words to learn was equivalent for the target and control conditions. In other words, if a child had 10 words to learn, that would mean 10 target and 10 control words were tracked. Total means the number of words produced using any of the data collection methods. Probe means the number of words produced in the probe condition. In-Tx means the number of words produced in treatment. Home means the number of words the child reportedly produced at home.

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Ranges for amount of input provided to each participant

Participant	Range of Input for Words Produced	Range of Targets Provided Per Minute
WL01	136 – 2091	4.85 – 14.67
WL02	216 – 584	8.27 – 10.09
WL03	257 – 2015	5.86 – 12.07
WL07	116 – 1492	7.3 – 9.63