

## Research Article

# Exploring Input Parameters in an Expressive Vocabulary Treatment With Late Talkers

Mary Alt,<sup>a</sup> Heidi M. Mettler,<sup>a</sup> Jessie A. Erikson,<sup>a</sup> Cecilia R. Figueroa,<sup>a</sup>  
Sarah E. Etters-Thomas,<sup>a</sup> Genesis D. Arizmendi,<sup>a</sup> and Trianna Oglivie<sup>a</sup>

**Purpose:** The aims of this study were (a) to assess the efficacy of the Vocabulary Acquisition and Usage for Late Talkers (VAULT) treatment and (b) to compare treatment outcomes for expressive vocabulary acquisition in late talkers in 2 conditions: 3 target words/90 doses per word per session versus 6 target words/45 doses per word per session.

**Method:** We ran the treatment protocol for 16 sessions with 24 primarily monolingual English-speaking late talkers. We calculated a *d* score for each child, compared treatment to control effect sizes, and assessed the number of words per week children acquired outside treatment. We compared treatment effect sizes of children in the condition of 3 target words/90 doses per word to those in the condition of 6 target words/45 doses per word. We used Bayesian repeated-

measures analysis of variance and Bayesian *t* tests to answer our condition-level questions.

**Results:** With an average treatment effect size of almost 1.0, VAULT was effective relative to the no-treatment condition. There were no differences between the different dose conditions.

**Discussion:** The VAULT protocol was an efficacious treatment that has the potential to increase the spoken vocabulary of late-talking toddlers and provides clinicians some flexibility in terms of number of words targeted and dose number, keeping in mind the interconnectedness of treatment parameters.

**Supplemental Material:** <https://doi.org/10.23641/asha.11593323>

In this study, we provided two versions of an expressive vocabulary treatment to late-talking toddlers. Our purpose was twofold: (a) to examine the efficacy of an expressive vocabulary treatment protocol first implemented by Alt, Meyers, Oglivie, Nicholas, and Arizmendi (2014) and (b) to compare treatment outcomes for expressive vocabulary acquisition in late talkers in two conditions. Children in one condition were assigned three different target words, which their clinician used in sentences, 90 times each per session. Children in the other condition were assigned six different target words, which their clinician used in sentences, 45 times each per session. Each time the clinician said a target word, it was counted as a single dose. This contrast allowed us to determine which dose number (45 or 90) per target word per session was more effective and to examine

how many target words (three or six) a clinician could successfully address in a given session. Discovering which parameters of treatment lead to the best outcomes can give clinicians clear principles to follow when implementing treatment. However, despite recent calls for better specification of treatment parameters (e.g., Hoffmann et al., 2014; Schulz, Altman, & Moher, 2010), these are not commonly specified in the language treatment literature.

## *Efficacy of the Vocabulary Acquisition and Usage for Late Talkers Protocol*

In the Vocabulary Acquisition and Usage for Late Talkers (VAULT) protocol, clinicians produce a set of target words at high frequencies in a variety of linguistic and physical contexts during play-based clinician-child interactions in which children are not required to produce language. The goal is to increase expressive vocabulary in late-talking toddlers. This input-based protocol utilizes cross-situational learning founded on principles of implicit learning. Alt et al. (2014) demonstrated the feasibility of this protocol (although it was not called VAULT at the time), but the efficacy of its different parameters remains unknown.

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

<sup>a</sup>Department of Speech, Language, and Hearing Sciences, The University of Arizona, Tucson

Correspondence to Mary Alt: [malt@email.arizona.edu](mailto:malt@email.arizona.edu)

Editor-in-Chief: Sean M. Redmond

Editor: Stephen M. Camarata

Received April 16, 2019

Revision received June 11, 2019

Accepted September 19, 2019

[https://doi.org/10.1044/2019\\_JSLHR-19-00219](https://doi.org/10.1044/2019_JSLHR-19-00219)

To define key treatment parameters, treatment research must be carried out in multiple phases. Fey and Finestack (2008) described a five-phase model of intervention research, starting with pretrial studies and followed by feasibility studies, early efficacy studies, later efficacy studies, and effectiveness studies. This study is an early efficacy study, the goal of which is to determine if potentially useful treatment variables exist in a cause-and-effect relationship with the measured treatment outcomes (Fey & Finestack, 2008). The objective of this study was to further investigate the viable treatment variables in the VAULT protocol, specifically number of treatment targets and dose number per target per session.

### Specifying and Defining Treatment Terminology

Tables 1 and 2 provide information about our treatment design following reporting guidelines (Campbell et al., 2018; Hoffman et al., 2014) and specifying and defining our treatment parameters. There are differences of opinion within the field concerning the use of terminology for treatment research. Our use of the terms *dose*, *dose rate*, and *number of treatment targets* differs from the previous literature in how they are used and/or reported. Table 2 defines these terms for our study, but a more detailed explanation can be found in Supplemental Material S1 provided with this article.

### Dose Rate

Although most vocabulary intervention studies have not reported dose rate, a handful have reported the components that allow it to be calculated. Table 3 outlines these components as reported by five word-learning studies with children between the ages of 20 months and 5 years. The optimal dose rate remains unknown, and the average rates of 1.55–5.16 doses per minute in these studies might fall below the optimal level for achieving the best treatment outcomes with late-talking toddlers. These studies reported the necessary information for calculating dose rate, but either they failed to report treatment effect sizes altogether or the effect sizes that they did report provide no insight relating to the effect of dose rate, one of the parameters we control for in the current study.

Evidence on word learning from the statistical learning literature suggests that input consisting of a high rate of target words can result in successful learning outcomes. Aslin, Saffran, and Newport (1998) presented 8-month-old infants with a continuous stream of four target nonwords repeated in a randomized order for 3 min with a dose rate across all target words of 90 words per minute. Despite the strikingly short period of exposure to target words, infants still demonstrated learning, which was likely feasible because of the high-rate nature of the input. Although replicating this extremely high dose rate would be difficult and

**Table 1.** The 12 steps of the TIDieR protocol for intervention description and replication.

TIDieR step	Item description	Page in manuscript where item is reported
1. Brief name	Provide the name or a phrase that describes the intervention.	1, 3
2. Why	Describe any rationale, theory, or goal of the elements essential to the intervention.	1–5
3. What (materials)	Describe any physical or informational materials used in the intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed (e.g., online appendix, URL).	3, 5, 7–9
4. What (procedures)	Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities.	3, 5, 7–11
5. Who provided	For each category of intervention provider (e.g., psychologist, nursing assistant), describe their expertise, background, and any specific training given.	3, 9
5a. Who received <sup>a</sup>	Describe the intended participants of the intervention.	1, 5–6
6. How	Describe the modes of delivery (such as face-to-face or by some other mechanism, such as Internet or telephone) or the intervention and whether it was provided individually or in a group.	3, 8–9
7. Where	Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features.	3, 7–8
8. When and how much	Describe the number of times the intervention was delivered and over what period including the number of sessions, their schedule, and their duration, intensity, or dose.	3, 7–9
9. Tailoring	If the intervention was planned to be personalized, titrated, or adapted, then describe what, why, when, and how.	5, 10
10. Modifications	If the intervention was modified during the course of the study, described the changes (what, why, when, and how).	Appendixes A and B
11. How well (planned)	If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them.	9–10
12. How well (actual)	If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned.	9–10

*Note.* Except where noted, text was taken directly from *The BMJ*, “Better Reporting of Interventions: Template for Intervention Description and Replication (TIDieR) Checklist and Guide,” Hoffman et al., Vol. 348, g1687, Copyright © 2014, with permission from BMJ Publishing Group Ltd.

<sup>a</sup>Not included in the original TIDieR checklists (Campbell et al., 2018; Hoffman et al., 2014).

**Table 2.** Treatment parameters.

Treatment parameter	Definition	Specification for current VAULT protocol
Dose	The action(s) in treatment that produce(s) change for a given target	The clinician's verbal model of the target word in a nontelegraphic, grammatical utterance
Dose number	The number of doses per target per session	Higher dose number condition: 90 teaching episodes (270 doses:3 words) Lower dose number condition: 45 teaching episodes (270 doses:6 words)
Total number of doses	The number of doses across all targets per session	270 doses per session for all target words combined
Dose rate	The total number of doses that occur within a given unit of time	Nine doses per minute (270 doses/30 min)
Dose form	The procedure used to administer the dose	Input-based, focused stimulation procedures using varied linguistic contexts, a variety of different activities, and no requirement for child productions
Treatment context	The setting or context in which the treatment is provided	In a child-friendly clinic room that includes the child, clinician, scorekeeper, and family members
Number of targets	The number of different targets that will be addressed in a given session	Three different target words versus six different target words
Session frequency	The daily, weekly, or monthly schedule of treatment sessions	Two times per week
Session duration	The duration of an individual session	30 min per session
Total intervention duration	The total number of days, weeks, or months the intervention is provided	16 sessions, with an average of two sessions per week, for an average of 8 weeks of treatment
Cumulative treatment intensity	Total Number of Doses × Session Frequency × Total Intervention Duration	270 × 2 × week × 8 weeks = 4,320

*Note.* The creation of this table was inspired by the work of Warren et al. (2007). While some of the terms and definitions are identical to theirs, others are new or modified to fit a broader range of treatment scenarios. VAULT = Vocabulary Acquisition and Usage for Late Talkers.

unnatural in a treatment setting, the infants' successful learning gives reason to think that increasing the dose rate in therapeutic contexts might also result in positive learning outcomes.

Alt et al. (2014) used a dose rate of 9.66 target words per minute across all target words combined in a word-learning feasibility study with late-talking toddlers. This rate aligns closely with Yu, Suanda, and Smith's (2017) findings that parents, when instructed to interact naturally with their 9-month-old child, named target items at a rate of 9.62 times per minute. In Alt et al.'s study, toddlers demonstrated gains on target words based on pre- to posttreatment measures, with an average effect size of 3.79 across all the late-talking toddlers. The toddlers also showed improvement on percentile rankings on a standardized parent report measure. These outcome measures suggest that high dose rates can lead to positive learning outcomes for late talkers.

Rice, Oetting, Marquis, Bode, and Pae's (1994) study also supports using a higher rate of input with late-talking toddlers: The children with language learning difficulties in their study demonstrated learning only at the higher dose rate (5.16 words per minute), but not at the lower rate (1.55 words per minute). MacRoy-Higgins and Montemarano (2016) recently suggested that late-talking toddlers could benefit from increased rates of exposure to target words in intervention. They found that allocation of attention was a significant predictor for word learning in both toddlers with and without language delays. Although it is not conclusively known whether late-talking toddlers have reduced attentional abilities, the finding that attention plays a significant role in word learning implies that a higher rate of exposure to target words could help compensate for attentional deficits that may influence their word learning. In part due to lack of specificity of terminology and details

**Table 3.** Word-learning studies with treatment parameters related to dose density.

Study	Number of target words	Teaching episodes per target word	Session length (min)	Dose density
Leonard et al. (1982)	16	5	45	1.78 words/min
Wilcox et al. (1991)	10	15	45	2.22 words/min
Rice et al. (1992)	10	6	12	5 words/min
Rice et al. (1994)	8	3	15.5	1.55 words/min
Rice et al. (1994)	8	10	15.5	5.16 words/min
Solomon-Rice & Soto (2014)	5	10	20	2.5 words/min
Solomon-Rice & Soto (2014)	10	10	20	5 words/min

reported in intervention studies, the optimal dose rate to use in word-learning treatment for late-talking toddlers has yet to be determined. However, there is reason to predict better learning outcomes for toddlers exposed to a higher rate of input.

Another reason to predict better learning outcomes for toddlers exposed to higher dose rates is related to a principle of statistical learning: regularity. The regularity principle (Plante & Gómez, 2018) states that the treatment target(s), which in our present study are the words targeted for learning, should be the most frequently and consistently presented input in the treatment session. In this study, we manipulated the dose number per target word per session (45 or 90), in essence changing the regularity of the target words between the conditions. In the 2014 feasibility study by Alt et al., clinicians provided input centering on a minimum of three target words per session. The target words occurred on a very regular basis: A clinician produced each target word a minimum of 64 times in a grammatical utterance within a short period (30–50 min). Regularity was enhanced through the high presentation rate of the target words during treatment sessions, making treatment targets the most consistently and frequently occurring input in the session. This study demonstrated that a relatively high rate of dose delivery was feasible: Not only were clinicians able to deliver the treatment with fidelity, but late-talking toddlers were able to tolerate and benefit from this treatment variable as evidenced by their positive word-learning outcomes. However, the relationship between the number of doses per target word and the word-learning outcomes was unclear. Although all children received a minimum of 64 doses per target word, the exact number of doses and dose rates per session varied and were not controlled across children and across each child's target words. In addition, the total number of sessions varied (14–20 sessions), as did the length of the sessions (30–50 min).

### Number of Treatment Targets

While dose rate is an important parameter to specify in treatment research, it alone does not answer a basic question about word-learning treatment: How many different words should I target per session? Clinicians often work on multiple targets and even multiple goals or objectives within a single session. The number of targets within a session interacts with the dose rate. Suppose a clinician wanted a dose rate of nine doses per minute for a 30-min session. This would mean that she would need to provide 270 doses for the 30-min session. However, if the clinician chose to focus those doses on a single target, with a ratio of 270:1, that would be more intense (i.e., 270 doses per target word) than if the doses were distributed across 10 different targets, with a ratio of 270:10, resulting in 27 doses per target word per session. Because these choices do not occur in isolation, there are consequences to either choice. We are not aware of any studies that have contrasted dose number per target word per session while keeping the overall dose rate constant, which we addressed in the current study.

## Late-Talking Toddlers

Late-talking toddlers are eligible to receive early intervention services under the Individuals with Disabilities Education Improvement Act of 2004. Late talkers are defined by a delay in expressive language, generally identified when a child is between 2 and 3 years of age (Desmarais, Sylvestre, Meyer, Bairati, & Rouleau, 2008), that cannot be explained by sensory/motor deficits, genetic or neurological disease, or another primary disorder (e.g., autism spectrum disorder; Desmarais et al., 2008; Rescorla, 2011; Singleton, 2018).<sup>1</sup> Specifically, late-talking toddlers demonstrate smaller expressive vocabularies and use fewer multiword utterances than their peers with typical language development. Although specific details of how this population is defined have varied across studies,<sup>2</sup> late-talking toddlers generally have a limited vocabulary and primarily produce single-word utterances, whereas typically developing 24-month-olds have up to several hundred words in their vocabulary and speak in simple two- to four-word sentences (e.g., Centers for Disease Control and Prevention, 2018).

### Language Treatment for Late Talkers

There is evidence that word-learning interventions are generally effective at improving expressive language outcomes for late-talking toddlers (DeVeney, Hagaman, & Bjornsen, 2017). A systematic review of word-learning interventions for late-talking toddlers found positive treatment effects across multiple studies for scores on formal language assessments, mean length of utterance, and the production of specific target words (Cable & Domsch, 2011). Although Cable and Domsch (2011) document a range of medium to large effect sizes across outcome measures and studies, treatment effect sizes for the production of specific target words were particularly high (Hedges's *g* values from 1.03 to 1.14). Regrettably, the specific treatment parameters of many of these studies have been underspecified, making them difficult to replicate or apply to treatment. For example, not all studies disclose the dose number per target word for

<sup>1</sup>Late-talking toddlers (i.e., "late talkers") have also been referred to by a variety of terms, including but not limited to children or toddlers with "late language emergence" (Cable & Domsch, 2011; Zubrick, Taylor, Rice, & Slegers, 2007), "expressive vocabulary delays" (Girolametto, Pearce, & Weitzman, 1996), "delayed language development" (Robertson & Ellis Weismer, 1999), and "specific expressive language disorder" (Buschmann, Multhauf, Hasselhorn, & Pietz, 2015).

<sup>2</sup>Some researchers have used age-based norms to determine cutoffs for late-talking toddlers (e.g., less than 10th percentile on a test of vocabulary), while others have used a cut-point for number of words spoken (e.g., 50 words or fewer for children 24–36 months old; see Desmarais et al., 2008, for a review). Some researchers have restricted their definition of late-talking toddlers to children who also demonstrate normal language comprehension. Other researchers have applied this term more broadly to toddlers with expressive deficits, regardless of whether or not they demonstrate normal receptive language skills (see Desmarais et al., 2008). Whichever metric is used, late-talking toddlers are clearly behind in expressive language development when one considers the vocabulary of a typical 24-month-old.



a session, so we do not know how many times researchers presented a new word to a learner.

### **Nature and Outcomes of Interventions**

Treatment approaches and techniques for late-talking toddlers have been primarily input based. Examples include focused stimulation (Girolametto, Pearce, & Weitzman, 1996) and modeling with or without imitation (DeVeney, Cress, & Reid, 2014; Ellis Weismer, Murray-Branch, & Miller, 1993). These treatments typically provide repeated exposures to target words in a child-friendly, play-based context. In some treatment procedures, clinicians also attempt to elicit productions from the child during treatment (DeVeney et al., 2014; Ellis Weismer et al., 1993). Input-based approaches have been found to improve expressive language skills in late-talking toddlers (Alt et al., 2014; DeVeney et al., 2014; ; Ellis Weismer et al., 1993; Girolametto et al., 1996), with or without output components. However, as discussed above, the current research does not allow clinicians to clearly specify how much input is optimal or necessary to achieve positive outcomes.

Although the outcomes associated with treatment for late talkers are promising, Cable and Domsch (2011) noted that the number of intervention studies was small, with 11 studies included and only four for which effect sizes could be calculated. Furthermore, Cable and Domsch reported that only four of 11 studies provided information on treatment fidelity checks. Therefore, as a research and clinical community, we are left with not only a limited number of treatment studies but also an insufficient amount of detail to replicate the existing treatment research. With expressive language delay in toddlers occurring at a prevalence of 10%–20% (American Speech-Language-Hearing Association, 2019), it is important to carry out treatment research to determine the parameters that promote success in treatments for late-talking toddlers.

### **The Current Study**

We defined dose as a clinician's single verbal model of a target vocabulary item in a nontelegraphic, grammatical utterance. Dose form was an input-based, focused stimulation procedure using varied linguistic contexts with a variety of activities per target word with no requirements for child productions. We controlled the number of sessions (16), the total number of doses per session (270), and the length of each session (30 min), yielding a dose rate of nine doses per minute. We compared the number of target words that were addressed within each 30-min session (three or six), with the goal of identifying the dose number per target word per session that would lead to stronger positive outcomes in toddler expressive vocabulary. Specifically, we contrasted a condition that had a higher dose number per target word (three target words/90 doses per target word per session) to a condition that had a lower dose number per target word (six target words/45 doses per target word per session). Both conditions adhere to the principle of regularity, but regularity is amplified in the three target/90 dose

condition. The current study includes a larger number of late-talking toddlers and uses a more tightly controlled VAULT protocol than that used in Alt et al. (2014). We predicted the following:

1. We would be able to replicate the findings from Alt et al. (2014) using the VAULT protocol and find evidence that treatment was more effective than no treatment for improving expressive vocabulary in late-talking toddlers.
2. The higher dose number per target word per session (three word/90 dose) condition would be more effective than the lower dose number per target word per session (six word/45 dose) condition.

## **Method**

### **Participants**

Participants included 24 primarily monolingual English-speaking late-talking toddlers (ages 25–41 months) whose families were willing to bring them to all evaluation, treatment, and follow-up appointments. Participants were quasirandomly assigned to either the three word/90 dose per target per session treatment condition (hereafter referred to as the higher dose number condition) or the six word/45 dose per target per session treatment condition (hereafter referred to as the lower dose number condition). If the child met all study criteria, he or she was assigned to either the higher dose number or the lower dose number treatment condition. New participants were matched to an existing participant if they were (a) the same sex and (b) within 3 months of age from each other. For each pair, the first participant was randomly assigned to either the higher or the lower dose number condition using a premade list of assignments generated from the website <https://www.random.org> (Random.org, n.d.). The corresponding match was assigned to the remaining condition. See Table 4 for demographic information for participants in each treatment condition.

Using institutional review board-approved materials, participants were recruited in the community with flyers distributed through places such as pediatricians' offices, libraries, and websites. Members of the research team also spoke with parents at events such as story times and "stay-and-play" events at public libraries. Children could not have other diagnoses (e.g., autism, hearing loss) and had to be primarily monolingual. Though most participants' parents reported speaking only English in the home, two participants (13 and 20) had some degree of exposure to another language (see Appendix A for modifications for these participants in treatment). However, the parents of these participants also reported English as the primary language in the home.

### **Inclusionary/Exclusionary Criteria**

Whether a child qualified for the study was determined across several stages prior to treatment. See Table 5

**Table 4.** Demographic information for participants by treatment condition.

Characteristic	Higher dose number condition	Lower dose number condition
<i>n</i>	15 (six females, nine males)	9 (five females, four males)
Age in months, <i>M</i> ( <i>SD</i> )	30.87 (5.52)	29.67 (4.79)
Standard score on the Bayley, <sup>a</sup> <i>M</i> ( <i>SD</i> )	93.21 (6.96)	97.22 (11.75)
Number of words produced on the initial MCDI, <sup>b</sup> <i>M</i> ( <i>SD</i> )	71.6 (139.45)	68.33 (93.32)
Number of words understood on the initial MCDI, <i>M</i> ( <i>SD</i> )	343.8 (146.48)	341.44 (131.36)
Race	11 White, 3 more than one race, 1 Black/African American	7 White, 2 more than one race
Ethnicity	8 Hispanic, 7 non-Hispanic	3 Hispanic, 6 non-Hispanic
Maternal education		
High school	1	
Associate degree or some college	6	3
Bachelor's degree	2	3
Graduate degree	6	3

*Note.* Bayley = Bayley Scales of Infant and Toddler Development—Third Edition (Bayley, 2006); MCDI = MacArthur–Bates Communicative Development Inventories: Words and Sentences (Fenson et al., 2007).

<sup>a</sup>Data are available for the Bayley for 23 participants. One participant in the higher dose number condition was outside the age range for this measure, and that participant's nonverbal intelligence was measured using the Kaufman Assessment Battery for Children—Second Edition (Kaufman & Kaufman, 2004). That child's standard score was 93. Testing for one child in each condition was stopped before reaching a ceiling, likely underestimating those children's actual abilities. <sup>b</sup>All participants but one fell below the 5th percentile based on the number of words produced for their age/sex on the MCDI. One participant's score was at the 5th percentile.

**Table 5.** Inclusionary/exclusionary criteria for treatment.

Stage	Purpose	Inclusionary criteria and measures
Before the evaluation	To rule out the presence of other diagnoses and influence of bilingualism on treatment outcomes	Parent phone interview <ul style="list-style-type: none"> <li>• Report no other diagnoses</li> <li>• Report of using primarily English in the home</li> </ul>
	To establish the presence of an expressive language delay and qualify for the evaluation phase	Below the 10th percentile <ul style="list-style-type: none"> <li>• on the MacArthur–Bates Communicative Development Inventories (MCDI): Words and Sentences (Fenson et al., 2007; if below 30 months old) or</li> <li>• on the MCDI-III (Dale, 2007; if 30 months or older)</li> </ul>
During the evaluation	To establish normal nonverbal intelligence	Standard score $\geq 75$ on the Bayley Scales of Infant and Toddler Development—Third Edition (Bayley, 2006) or on the Kaufman Assessment Battery for Children—Second Edition (Kaufman & Kaufman, 2004) for one participant who exceeded the age limits of the Bayley <ul style="list-style-type: none"> <li>• Pass near-vision screening at 20/40</li> <li>• Pass hearing screening using play audiometry at 20 dB for 1000, 2000, and 4000 Hz</li> </ul>
	To establish normal vision and hearing	If screening could not be completed, participants received “functional pass” if the parents and the evaluators had no concerns based on the child's behavior responding to visual and auditory stimuli during the evaluation.
Delayed start condition only Following 8-week delay prior to treatment	To establish the presence of an expressive language delay and rule out improvement due to maturation	Below the 10th percentile <ul style="list-style-type: none"> <li>• on the MacArthur–Bates Communicative Development Inventories (MCDI): Words and Sentences (Fenson et al., 2007; if below 30 months old) or</li> <li>• on the MCDI-III (Dale, 2007; if 30 months or older)</li> </ul>

*Note.* Due to the attentional limitations of this age group, we did not always reach a ceiling. For two participants (12 and 19), once we reached a threshold that ruled out intellectual disability (standard score of  $\geq 75$ ), we discontinued the test.

for a description of all inclusionary/exclusionary criteria. Before the evaluation, all families completed the MacArthur–Bates Communicative Development Inventories–Second Edition (MCDI; Fenson et al., 2007), which provided information about their child’s current vocabulary and language skills. If a child scored in the 10th percentile or below on the MCDI, the family was invited for an evaluation. All participants scored at or below this cut-point. One participant was excluded at this point in the study because we were unable to contact the family for an evaluation.

Normal nonverbal intelligence, hearing, and vision were established during the evaluation. All participants demonstrated nonverbal abilities within the normal range, and we did not exclude any children from the study based on vision or hearing results. However, one participant withdrew from the study after beginning treatment because he needed to have surgery to insert pressure equalizing tubes.

A subset of participants had a delayed start to ensure that gains made in treatment were not due to only time and maturation. The parents of these toddlers filled out the MCDI a second time prior to beginning treatment. If a participant’s MCDI score placed them above the 10th percentile at this point, it was assumed that the child was improving based on maturation and was no longer eligible for the study. None of our participants was excluded based on their post-delay MCDI percentile.

## **Materials and Procedure**

### **Pretreatment**

Once a family contacted us, we conducted a phone screening to see if they were likely to be eligible. If they passed this initial screening, we sent them a consent form and the MCDI. If the child met our criteria for the MCDI, we conducted a structured parent interview to obtain further information about the child (e.g., why the family thought their child could benefit from the study, if the child had any other diagnoses). If the responses during this interview aligned with inclusionary criteria, the family was invited to the clinic for a full evaluation in which we administered a nonverbal intelligence test and screenings of hearing and vision.

All evaluations took place at the Grunewald-Blitz Center for Children’s Communication Disorders, the pediatric clinic in the University of Arizona’s Speech, Language, and Hearing clinics in Tucson, Arizona. The evaluation sessions lasted approximately 60–75 min, which was largely dependent on the individual child’s tolerance of and participation in the assessment activities. Due to the attentional limitations of this age group, we did not always reach a ceiling on the test of nonverbal intelligence. For two participants (12 and 19), once we reached a threshold that ruled out intellectual disability (standard score of  $\geq 75$ ), we discontinued the test. For another participant (15), we accepted scores on a nonverbal intelligence test from the Department of Economic Security statewide program responsible for

early identification (Arizona Early Intervention Program) in lieu of retesting that participant.

We attempted to screen participants’ near-vision and hearing acuity to ensure that they would be able to interact with treatment materials and clinicians. With this young population, obtaining reliable responses was challenging and not necessarily feasible. For the vision screening, we used a near-vision examination card with shapes and separate cards with larger versions of the corresponding shapes. We asked the participants to match the shapes on the cards with the shapes on the 20/40 line. For the hearing screening, we used play audiometry. Most children lost interest after the training, did not tolerate wearing the tight-fitting supra-aural headphones, or both. For both vision and hearing, we also requested any medical records for recent vision and hearing evaluations and used these to confirm vision/hearing status. If we had any concerns that children might have had compromised sensory systems, we delayed treatment until a formal audiological evaluation could be completed. This was the case for Participants 4 and 12.

If children were in the delayed start ( $n = 14$ ), we scheduled an appointment roughly two months in the future ( $M = 7.83$  weeks), at which point the baseline for target/control words was established. The delayed start allowed us to measure growth without intervention, in order to better interpret treatment outcomes compared to maturation. If children began treatment immediately with no delay ( $n = 10$ ), we scheduled baseline sessions until we were certain the child did not produce any of the target or control words. Assignment to the delayed/immediate start groups was not randomized. We tried to have as many participants as possible in the delayed start group in order to have a control for maturation, but staffing and scheduling constraints meant some children began treatment immediately. Bayesian  $t$  tests showed anecdotal evidence for no difference between the delayed start and immediate start groups in terms of age ( $BF_{10} = 0.492$ ) and nonverbal intelligence ( $BF_{10} = 0.740$ ). The sex distribution was equivalent for the delayed start (nine boys, five girls; 64.3%/35.7%) and immediate start (six boys, four girls; 60%/40%) groups.

### **Baseline Sessions**

Just prior to treatment, we verified that children were not using either the target or the control words by having them participate in at least three baseline sessions on three separate days. During these sessions, the children were shown pictures of potential target and control words and asked to produce the words. The pictures were presented on a Samsung Galaxy Tab A tablet. For most nouns, the examiner presented the picture and asked, “What is this?” For most verbs, the examiner asked, “What is he/she doing?” Other types of words were presented with a cloze phrase such as “This pig is clean, but this pig is \_\_\_\_\_.” If the child produced any of the words during the baseline sessions, or if a parent reported that the child said that word at home, the word was replaced by a different word, and the child was retested until a 3-day baseline was established for each

target and control word. The images used in the baseline probe sessions were not used during treatment.

### Stimuli and Target Selection

Each child had 10 pairs of words (10 targets and 10 controls) that were selected from the MCDI and matched to be as equivalent as possible. Whenever possible, we included words that the family reported they would like their child to use. We asked parents to fill out a form by listing up to 50 preferred words. In order for a word to be a viable candidate for treatment/control, it had to be a word that the child was not using, but did understand, according to parent report. One potential candidate was excluded from the study because the parents could not identify any words that the child understood but did not say. Parents filled out a modified version of the MCDI (printed on pink paper, with the word RECEPTIVE printed on it), where they were asked to indicate which words their child understood.

In order to choose word pairs (i.e., target/control), we looked for words that the child understood but did not say that (a) were from the same grammatical category (e.g., nouns, verbs, adjectives), (b) had similar semantic categories (we often started looking for pairs within the MCDI categorical groupings [e.g., animals, small household items]), (c) had the same number of syllables, and (d) had similar item trajectories on the Stanford Wordbank Item Trajectories ([http://wordbank.stanford.edu/analyses?name=item\\_trajectories](http://wordbank.stanford.edu/analyses?name=item_trajectories); Frank, Braginsky, Yuovsky, & Marchman, 2017), with a focus on similarities at the child's specific age. These trajectories show what percentage of children at a particular age—from 16 to 30 months—say a given word. For example, at 20 months of age, 50% of children say the word “horse.” Figure 1 provides three examples of word pairs and their item trajectories, generated from Wordbank.

Once 10 pairs of words were selected, we used pre-generated randomized lists to determine which word in the pair would be the target or control and in which order the words would be taught. We pregenerated multiple lists, so each participant's words were randomized according to a unique order.

The selected words were from different classes (e.g., nouns, verbs), but there was no formula for choosing a set

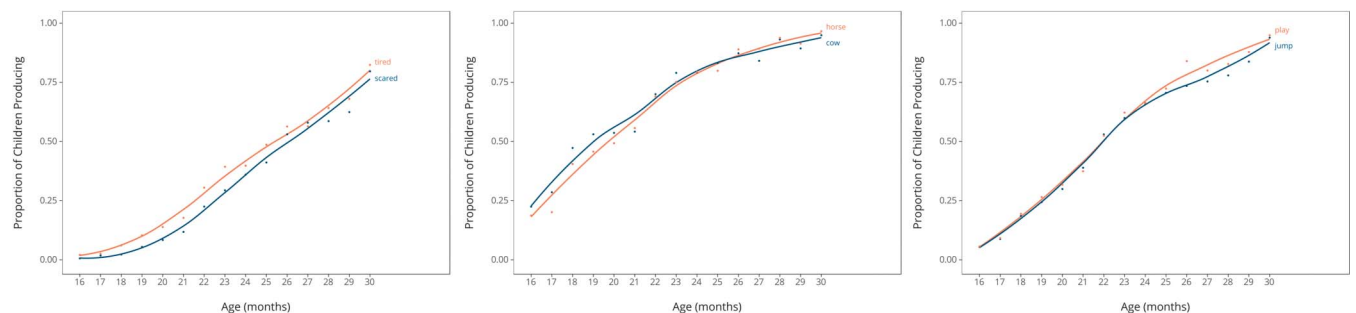
number of word classes. Of the 24 participants, 24 (100%) had nouns as part of their target words, 21 (87.5%) had verbs, 16 (66.6%) included adjectives, five (20.8%) had prepositions, and four (16.6%) had some other type of word. The average distribution of word classes across the 10 target words per participant, for both the higher dose number and lower dose number conditions, was six nouns, two verbs, and two adjectives.

### Treatment

After baselines were established, we began treatment. Children were seen at the university clinic for 16 treatment sessions lasting 30 min, generally twice a week for 8 weeks, for a total intervention duration of 480 min. All participants attended all 16 sessions. As noted above, the dose was a clinician's verbal model of a target vocabulary item in a non-telegraphic, grammatical utterance. The dose form was an input-based, focused stimulation procedure using varied linguistic contexts, with a variety of activities per target word with no requirements for child productions. The dose context was a child-friendly clinic room that included the child, clinician, scorekeeper, and family members. Both treatment conditions had an equal total number of doses: 270 per session. For a 30-min session, that resulted in a rate of nine doses per minute. The treatment conditions differed in number of targets and in dose number per target word per session. In the higher dose number condition, we targeted only three words per session, but in the lower dose number condition, we targeted six words per session. The cumulative treatment dose was 270 doses per session  $\times$  2 sessions per week  $\times$  8 weeks = 4,320 for both conditions.

A target word was addressed in treatment until the toddler said it at least one time per session across three consecutive sessions. Once a toddler met these requirements, a new target-and-control word pair was substituted. In some cases, toddlers heard the same three target words across all 16 sessions; these toddlers never produced their initial set of target words in three consecutive sessions. Other toddlers, however, met the requirements for multiple target words and so were exposed to a greater number of unique target words across their 16 treatment sessions. In cases when a toddler met our production requirements for all

**Figure 1.** Example of word trajectories for target-and-control word pairs.





10 target words, clinicians returned to the top of the target/control word pair list and began to re-address the same target words as they had earlier in the intervention. This scenario happened infrequently (two subjects).

Therapy activities were designed to allow for the natural use of the target words and the incorporation of examples of the control words. In practice, the treatment sessions looked like play. For example, if a child in the higher dose number condition was exposed to the words “milk,” “mom,” and “go,” a clinician might design an activity such as grocery shopping or preparing cereal. Although we did not use the labels of the control words, multiple representations of the control word items were included in each session and made salient in order to provide the child with opportunities to produce the control words. Often, clinicians referred to the control items using general vocabulary (e.g., “this,” “that”) or using synonyms (e.g., “have fun with” instead of “play”). So, in the example above, if the control words were “water,” “dad,” and “play,” in addition to mom pouring milk, dad might be playing with water.

Additionally, a variety of object exemplars for each word were presented during each session in order to ensure semantic variability and to keep children’s attention. For example, for milk, one could select a small toy milk bottle, a coloring page of a box of milk, a real gallon of milk, a printed picture of a glass of milk, and a children’s book that features milk. New activity stimuli were selected or created after every session in order to maintain variability of items across sessions.

Each participant had two clinicians to encourage talker variability and to lessen the potential of a clinician effect. Each clinician provided treatment 1 day each week. Clinicians had various levels of clinical experience ranging from experienced, licensed, certified speech-language pathologists to undergraduate students for whom this was their first clinical experience. All clinicians demonstrated fidelity to the VAULT protocol. Clinicians were trained to perform the VAULT treatment protocol by watching video recordings of ideal delivery of the treatment, which highlighted the overall treatment approach and took into account use of the target words and rate of delivery. Each clinician also met with a licensed, certified speech-language pathologist to ensure understanding of the protocol.

### **Posttreatment**

Following treatment, we had two posttreatment assessments. Parents also took part in a posttreatment interview. The first posttest was scheduled for the first day following treatment (or as soon as possible); the second posttest was scheduled 4–6 weeks after treatment. Families filled out the MCDI prior to both of those sessions. We also repeated the probes for target/control words at both of those appointments. We administered additional descriptive measures, if children were of an appropriate age, such as the Expressive Vocabulary Test–Second Edition (Williams, 2007), the Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000), and the MCDI-III (Dale, 2007). Posttreatment sessions were conducted by a lead member of the research team who was not the child’s

clinician. The structure of these sessions differed from treatment sessions in two ways. First, toddlers were explicitly asked to produce words; this confrontation-naming style was not part of our treatment protocol. The researcher simply directed the child’s attention to the tablet and asked, “What is this?” or “What is he doing?” No cues (e.g., phonological) were provided to the toddlers, and no references were made to materials used during treatment. Second, words were elicited by showing toddlers pictures on a tablet computer rather than by interacting with toys in a play-based manner, the latter of which characterized all treatment sessions. All posttreatment sessions took place in the same clinic as the treatment sessions. The specific room in which posttreatment sessions were held depended upon scheduling constraints and so did not systematically vary from the rooms used for treatment for each toddler. A research staff member who had not been involved with the treatment sessions conducted the posttreatment interview with the parent(s) or family members who had brought the child to the treatment sessions. Like the pretreatment interview, questions were open-ended and allowed for families to share their personal experience and impressions of the research study. Families were counseled about next steps for their child regarding treatment options in the community.

### ***Fidelity and Reliability***

During treatment, we ensured fidelity within the session (i.e., either 45 or 90 doses per target word) by using scorekeepers. Scorekeepers were present for nearly every session. During each session, a scorekeeper sat in during the session and tallied the number of times that each target and control word was produced. We asked family members to refrain from talking during the session so the doses could be controlled by the clinician. However, any production of the word from any person other than the child counted as a dose. The scorekeeper helped keep the clinician on track by discreetly signaling when the clinician reached the halfway mark for a word and letting the clinician know when she had reached the target number of doses for a word. If the clinician continued to produce the target words after the criterion had been reached, they were alerted by the scorekeeper.

Fidelity was calculated by comparing the actual number of doses provided per word within the 30-min session to the targeted number of doses. Overall, fidelity was 99.12% with a range of 97.36%–99.83%, with an average fidelity for the higher dose number condition of 99.35% and an average fidelity of 98.72% for the lower dose number condition. Though it was expected that clinicians would not use any control words, they were occasionally spoken during the session, either by the clinician, a sibling, a parent, or another adult in the room. Productions of control words were tracked, and clinicians or reliability trackers (described below) were alerted if they were using a control word repeatedly. The average number of times control words were said across all 16 sessions was 49.70, or 1.15% of the cumulative treatment intensity for the target words, with a range of 10–135. The average for the higher dose number condition

was 56.28 (1.27% of the cumulative treatment intensity for the target words), and the average for the lower dose number condition was 40.5 (0.91% of the cumulative treatment intensity for the target words).

One of our within-treatment outcome measures was a child's production of words, so we needed to ensure that we had a reliable measure of what the children said. Considering that our participants were toddlers, many verbal productions did not have adultlike forms. Clinicians were trained to write down any words or word approximations accompanied by unambiguous referents that the child used during the session. Protowords were not recorded. For the majority of sessions, a licensed, certified speech-language pathologist served as a reliability tracker. The reliability tracker was in the room during each session and listened to and wrote down the child's verbal productions. Although all therapy sessions were video-recorded for later review, we did not rely primarily on the videos to transcribe children's productions because they did not always capture a clear view of a child's articulators or contextual information such as referents (e.g., an image in a book).

The protocol for any verbal production was for the clinician and the reliability tracker to verbally confirm immediately after the utterance to see if they agreed upon the interpretation of the utterance. This typically involved either the clinician or the reliability tracker repeating the child's utterance verbatim while the other confirmed whether she agreed. When they disagreed or when a child's utterances were unintelligible or unclear, the clinician or reliability tracker noted the time so utterances could be reviewed on videotape following the session.<sup>3</sup> Only productions on which at least two team members agreed were counted as word productions. Children were generally engaged in the treatment activities and unbothered by these brief adult interactions. We also conferred with parents to help interpret utterances, but their judgments alone were never used to determine a child's production of a word. At the end of the therapy session, both parties reviewed their summaries of all productions and calculated reliability. To calculate reliability, we divided the total number of unique agreed-upon utterances by the total number of unique utterances produced, yielding a percent agreement. Utterances the child produced multiple times were only counted once for the purposes of calculating reliability. For example, if the child said "mama" 10 times, only one of those productions was included in the reliability calculation. Multiword utterances were treated holistically for reliability. For example, if one person thought the child said, "mama go," and the other person thought the child said, "my boat," it only counted as one disagreement toward reliability. Utterances that needed to be reviewed on video, and thus were unclear to both listeners, were not included in the reliability calculations. We were able to calculate reliability for 331 of 368 sessions

(89.94%, not including Participant 10; see Appendix B). Average reliability was 95% for both the higher and lower dose number conditions. The range was 92%–100% for the lower dose number condition and 87%–100% for the higher dose number condition.

### *Analytic Plan*

We had several dependent variables based on measures derived from within treatment sessions or from outside treatment. Our primary dependent variable, a within-treatment measure, was treatment effect size ( $d$ ), which is an effect size used in treatment research to measure learning for single subjects (Beeson & Robey, 2006). We calculated  $d$  by subtracting the mean of the first 3 days of treatment from the mean of the last 3 days of treatment and dividing by the standard deviation of the last 3 days of treatment. If there was no variability in those last 3 days, we used the smallest possible standard deviation (0.577). This protocol has been modified for treatment research with children from the standard protocol used by Beeson and Robey (2006) for single-subject design to use the standard deviation of the last three data points, rather than the first three data points, to account for the fact that children often demonstrate little to no variability at baseline for targets as they are often at zero. In order to calculate the mean, we counted how many of the target and control words were produced. Each word counted only one time. Thus, if a child said "milk" four times, it only counted as one word. A mean of 4 corresponds to a child producing four unique words at least one time each during a session.

Our secondary dependent variable was rate of growth on the MCDI, an outside-treatment measure, which we calculated as total number of words learned on the MCDI divided by number of weeks between MCDI administrations. We had a number of other within- and outside-treatment measures that we used to capture vocabulary growth: probes of the target and control words posttreatment and after follow-up, MCDI reporting of target/control words posttreatment and after follow-up, number of words treated, number of times target/control words were said during treatment, number of times words were said at least three times during treatment, and first session in which target/control words were said.

We used a combination of single-subject and group techniques to analyze these data, following protocols of treatment research (Alt et al., 2014; Plante et al., 2014; Plante, Tucci, Nicholas, Arizmendi, & Vance, 2018). Specifically, we calculated individual treatment effect sizes for each participant using  $d$ , which we then combined to make group-level comparisons. We used Bayesian repeated-measures analysis of variance (ANOVA) to determine if there were differences between the target and control words and if there were interactions between word type (i.e., target vs. control) and treatment condition (i.e., higher vs. lower dose number). We also used Bayesian  $t$  tests to examine other specific between-condition comparisons. Bayesian analysis techniques have several advantages that made them well

<sup>3</sup>Reliability procedures for Participant 10 were different from other participants because of the volume of output produced during each session. Please see Appendix B for details.

sued for this study. First, Bayesian analysis can manage differences in group sizes (Kruschke, 2013). Second, with Bayesian statistics, one can interpret not only between-group differences but also the probability of a lack of between-group differences (Kruschke, 2013). We used JASP (Version 0.9.1.0; JASP Team, 2018) for our data analyses. For the Bayesian *t* tests, we used JASP's default prior probability of a zero-centered Cauchy distribution set at 0.707, which is similar to a normal distribution, and for the Bayesian repeated-measures analysis of covariance, the *r* scale for fixed effects was .5 (Wagenmakers et al., 2018). While there are theoretical arguments for the use of differing priors, the benefits of the JASP defaults, including invariance to changes measurement scale, have been argued in Rouder, Morey, Speckman, and Province (2012) and Wagenmakers et al. (2018). The qualitative interpretation of Bayes factors (e.g., anecdotal, moderate, strong, extreme) is taken from Wagenmakers et al.

## Results

### *Is VAULT More Effective Than No Treatment?*

In order to answer this question, we compared the effect sizes for treatment and control words across treatment conditions. There was very strong evidence ( $BF_{INC} = 36.419$ ) that the treatment effect size ( $M = 0.99$ ,  $SD = 1.23$ ) was larger than the control effect size ( $M = 0.16$ ,  $SD = 0.43$ ). Fifteen of the 24 participants had an effect size larger than zero. See Supplemental Material S2 for graphs depicting individual performance on target and control words. Of the seven additional descriptive measures we used to characterize treatment, six of them showed at least moderate evidence that supports the idea that the VAULT condition was more effective than no treatment (see Table 6).

A second metric was whether the rate of words learned on the MCDI would be higher after treatment, compared to a delay with no treatment. We had 14 participants for whom treatment was delayed, and we tested their rates of word learning using a Bayesian paired-samples *t* test. There was moderate evidence ( $BF_{10} = 4.53$ ) of a higher rate of words learned on the MCDI posttreatment ( $M = 7.62$ ,  $SD = 8.21$ ) than the rate of words learned during the delay period ( $M = 2.77$ ,  $SD = 4.15$ ). Of the 14 participants with a delayed start, 12 had higher rates of words learned posttreatment than during the delay (see Figure 2). These data provide evidence that growth was more likely a result of treatment than of maturation.

However, we were also interested in rate of words learned outside treatment to see if children had “learned how to learn.” The group as a whole (delayed start and immediate start), on average, added 6.14 ( $SD = 7.29$ ) words per week on the MCDI posttreatment and 7.90 ( $SD = 9.29$ ) words per week at follow-up. While this is encouraging, examining individual data (see Figure 3) shows that just over half of the children (post,  $n = 15$ ; follow-up,  $n = 11$ ) made relatively modest gains of less than five words per week and the rest ( $n = 9$ ) made more impressive gains,

with an average of 13.68 ( $SD = 6.66$ ) words gained per week postintervention and 15.13 ( $SD = 8.73$ ) words gained per week at follow-up.<sup>4</sup> Ganger and Brent (2004) define a vocabulary spurt as learning roughly five words per week, and although there is a range of rates of learning in the literature, a rate of five words per week falls within the bound of what many researchers have found for “fast” learners (see Alt et al., 2014). Thus, only a subset of participants appeared to “learn to learn.”

### *Which Target Dose Number Was More Effective?*

We used a Bayesian repeated-measures ANOVA with word type (target vs. control) as the within-subject measure and condition (higher dose number vs. lower dose number) as the between-group measure to answer this question. First, we checked to see if there were any group differences related to age at the time of treatment. There was anecdotal evidence ( $BF_{10} = 0.42$ ) for the null hypothesis that there was no difference between the conditions. Because the mean ages for children in the two conditions differed by less than 1 month (see Table 4), we did not covary age. The results from the ANOVA provided anecdotal evidence ( $BF_{INC} = 0.877$ ) for no effect of condition. There was also anecdotal evidence ( $BF_{INC} = 0.587$ ) for no interaction between condition and word type. Figure 4 illustrates the effect sizes and standard errors for each group. Next, we examined the number of words learned per week at pre-treatment, posttreatment, and follow-up on the MCDI by condition, for the delayed start group. There was anecdotal evidence for no difference between conditions ( $BF_{INC} = 0.716$ ) for no interaction ( $BF_{INC} = 0.46$ ). Figure 5 shows these data. Then, we examined each of the seven outcome measures reported in Table 6 by condition. In every instance, there was anecdotal to moderate support for the null hypothesis ( $BF_{INC}$  ranged from 0.21 to 0.83). That is, it was more likely that treatment condition did not contribute to the outcome of the treatment.

## Discussion

Our results support the conclusion that the VAULT protocol described above is an efficacious treatment for the majority of late talkers in our study. In a relatively short amount of time (sixteen 30-min sessions, typically spread over 8 weeks), most children began producing targeted words, and there was moderate evidence that children with the delayed start improved their rate of word learning outside treatment, as measured by the MCDI, an outside-treatment measure. For the entire group, at least a third of them appeared to “learn how to learn,” gaining, on average, more than 13 new words per week on the MCDI, which is more than twice as many words as the “fast” learners noted in the literature described earlier (e.g., Alt et al., 2014;

<sup>4</sup>We did not obtain follow-up MCDI scores for three children. We were unable to schedule with two families, and the third family had moved.

**Table 6.** Comparison of target and control words on seven different metrics.

Metric	Target, <i>M (SD)</i>	Control, <i>M (SD)</i>	Evidence	Interpretation
Probe posttreatment <sup>a</sup>	1.54 (1.84)	1.04 (1.68)	BF <sub>10</sub> = 0.98	Anecdotal support for no difference between treatment and control words
Probe follow-up <sup>a</sup>	2.59 (2.36)	1.72 (1.80)	BF <sub>10</sub> = 5.81	Moderate support for target > control
Specific words MCDI posttreatment <sup>a</sup>	3.58 (2.63)	2.54 (2.90)	BF <sub>10</sub> = 5.92	Moderate support for target > control
Specific words MCDI follow-up <sup>a</sup>	4.47 (3.17)	3.28 (2.86)	BF <sub>10</sub> = 3.67	Moderate support for target > control
Number of times words said during treatment <sup>b</sup>	33.34 (55.85)	2.26 (5.84)	BF <sub>10</sub> = 8.33	Moderate support for target > control
Number of words said at least three times during treatment <sup>b</sup>	3.20 (2.84)	0.54 (0.72)	BF <sub>10</sub> = 3013.90	Extreme support for target > control
First treatment session in which words were said <sup>b</sup>	2.66 (3.57)	8.11 (3.44)	BF <sub>10</sub> = 15.47	Strong support for target > control

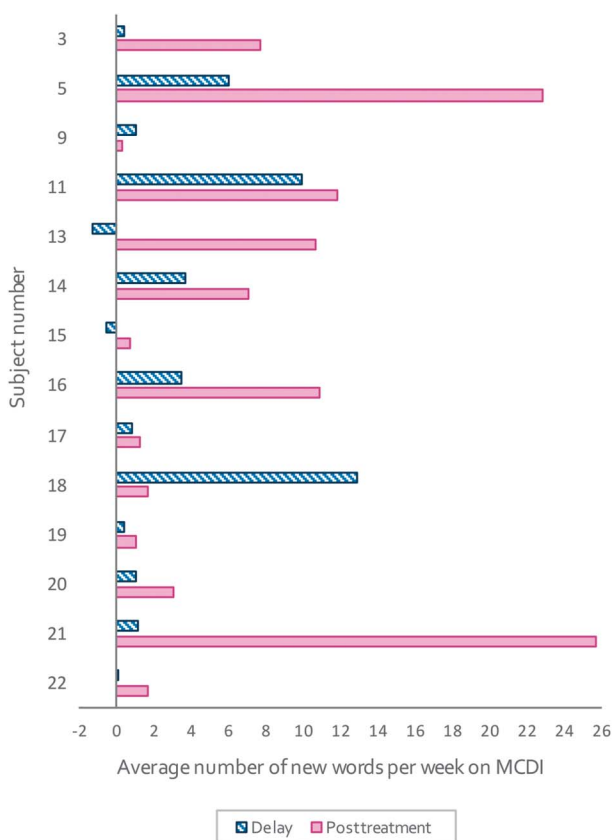
<sup>a</sup>Outside-treatment variable. <sup>b</sup>Within-treatment variable.

Ganger & Brent, 2004). Thus, there appears to be differential responses to the statistical learning aspect of the treatment, in which children implicitly learn rules—in this case, learning what to attend to in order to learn new words—without explicit instruction.

Our large treatment effect size, a within-treatment outcome measure, is in line with the four studies in Cable and

Domsch's (2011) review that provided data for effect sizes. In comparison with the studies reported by Cable and Domsch, we were able to achieve comparable results in a shorter amount of time (8 weeks of treatment compared to a range of 10 weeks to 6 months). What we have been able to add to the literature are detailed data on our treatment parameters, which will allow for clear comparisons of additional studies.

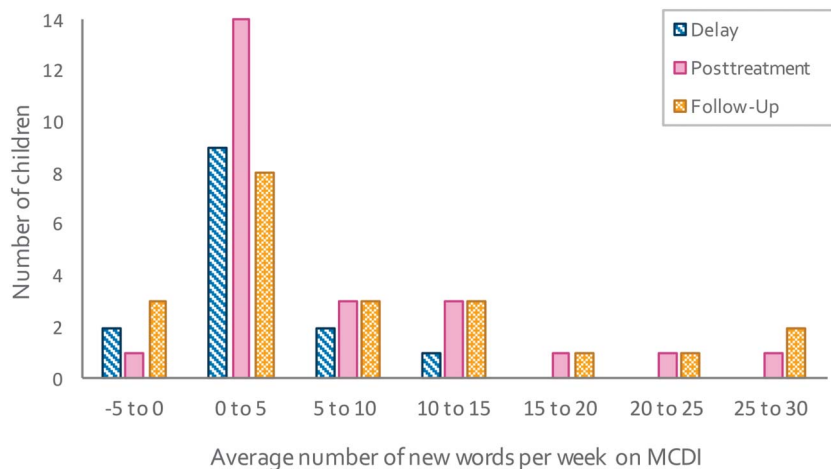
**Figure 2.** Average number of words gained during pretreatment and posttreatment for children with a delayed start. MCDI = MacArthur-Bates Communicative Development Inventories.



Support for efficacy of the VAULT protocol is based on evidence from both within-treatment (e.g., treatment effect size *d*, number of times words were said during treatment) and outside-treatment (e.g., posttreatment probe, rate of growth on the MCDI at follow-up) outcome measures. We acknowledge that evidence from within-treatment measures and performance on targeted words differ from those derived from outside-treatment measure or measures of generalization (i.e., learning to learn). Outcome measures that are farther removed from the treatment context may be interpreted as stronger indicators of the efficacy of treatment. However, we do not consider the inclusion of more proximal measures or within-treatment measures to be a limitation of the current study. Based on Fey and Finestack's (2008) model for phases of treatment research, the current study is an example of early efficacy treatment research. Because the major goal in this phase of treatment research is to establish early evidence for a cause-effect relationship, it is appropriate for outcome variables to be more targeted and proximal in nature, like the within-treatment measures in the current study (Fey & Finestack, 2008). These measures provide evidence of whether there was an effect of treatment. Outside-treatment measures, however, provide more information about how the treatment effect generalizes to other contexts, as well as other potential targets (e.g., words not explicitly targeted during therapy). This type of treatment measure is characteristic of later efficacy research. Because this is an early efficacy study, within-treatment measures play an important role in establishing a cause-effect relationship between the VAULT protocol and word learning. We included outside-treatment measures in order to gain early insight into VAULT's potential for more generalized effects of treatment. Although the effects for within-treatment measures were stronger than for outside-treatment measures,



**Figure 3.** Average number of words gained per week on the MacArthur–Bates Communicative Development Inventories (MCDI) posttreatment and at follow-up for all participants.

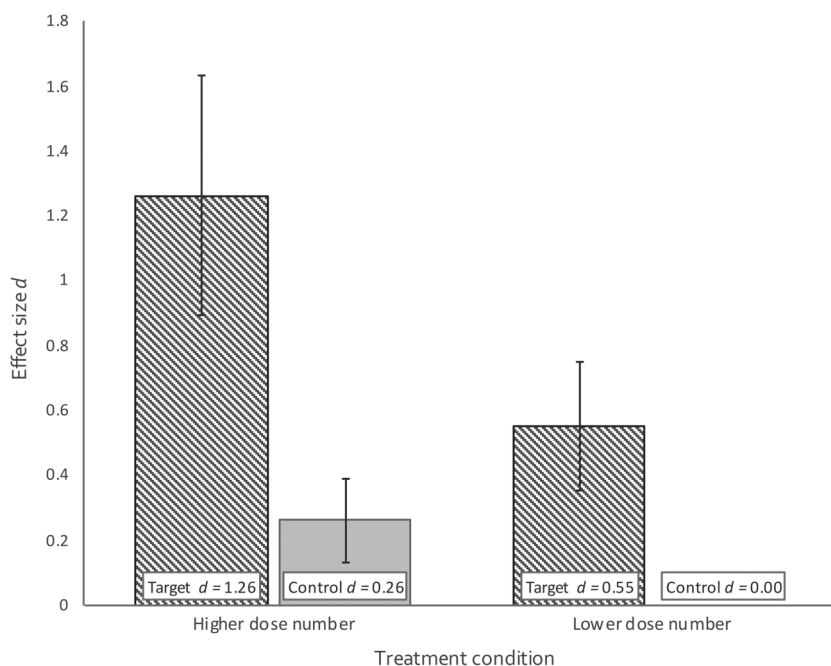


the evidence of effects for both types of measures suggests that VAULT is an efficacious treatment.

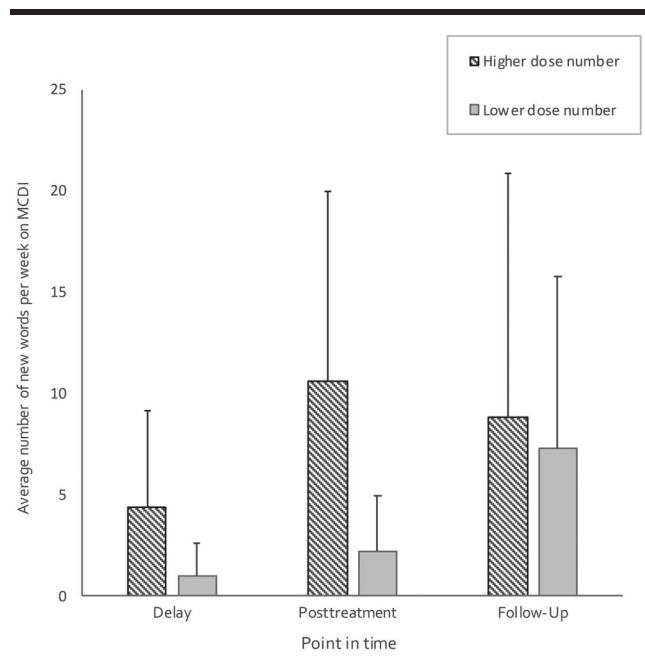
In terms of treatment parameters, we were interested in examining the role of dose number per target word and number of target words per session. We found anecdotal evidence favoring the hypothesis that there was no clear difference between the higher dose number (three target words/90 doses per word per 30-min session) and lower dose number (six target words/45 doses per word per 30-min session)

conditions. Importantly, there was evidence that the treatment worked for both the higher and lower dose number conditions. At first glance, one might interpret this finding as a recommendation for using the lower dose number condition, as it appears to be a more efficient use of time (targeting six vs. three words). However, we view this finding in a broader context and interpret it as license for clinicians to customize the number of words they teach based upon individual client characteristics, using an evidence-based range

**Figure 4.** Effect sizes and standard error by word type by condition. Higher dose number refers to the three word/90 dose number per word per session condition, and lower dose number refers to the six word/45 dose number per word per session condition.



**Figure 5.** Average number of words gained per week on the MacArthur–Bates Communicative Development Inventories (MCDI) with standard deviations at pretreatment, posttreatment, and follow-up by condition. Higher dose number refers to the three word/90 dose number per word per session condition, and lower dose number refers to the six word/45 dose number per word per session condition.



of dose number per target word. For example, a child with a shorter attention span might prefer the rapidly changing activities characteristic of the lower dose number condition, while a child who has difficulties with transitions might find this same condition too frenetic.

Selecting treatment parameters is a difficult and dynamic process. Each parameter potentially affects the others. See Figure 6 for a representation of the relationship between dose number and other treatment parameters. Our research question was “How many words? How many times?” Unfortunately, many clinicians never ask the “How many times?” question. However, one cannot answer the question “How many words?” without considering dose number, dose rate, and other treatment parameters. We have to be ready to take multiple factors into account. The answer to the question “How many words can I teach?” will never be simply “3” or “10.” The answer will be “3” (or 10), given a dose rate of  $X$ , a dose number per word of  $Y$ , and  $Z$  available minutes. For example, in this study, we found a positive treatment effect for some children by focusing on six target words. However, it is not just six target words; it is six target words with 45 doses per word per session at a rate of nine doses overall per minute in a 30-min session. It might be justified for a clinician to suppose that, if we could successfully teach six words, we might just as easily teach eight, increasing efficiency. However, if the clinician tried to teach eight words in the same time frame, using the same dose rate, the dose number per word would

decrease to just over 33. While this might work, we do not currently have data to support that dose number. Alternatively, if someone wanted to target eight words per session and keep the dose rate (nine words per minute) and the dose number per target word per session (45 doses) consistent with the higher dose rate condition, they would need to extend the time of their session to 40 min, which could be too long for a toddler.

So how could a clinician decide how many words to teach in a given session? She or he would need to make evidence-based decisions about dose number per target word per session, number of available sessions, time available within a session for vocabulary instruction, and dose rate. While the question can be answered using the equation in Figure 6, Figure 7 provides a friendlier flowchart based on data from this study and links to the terms specified in Figure 6. There are clearly many options for providing treatment, but we should be making decisions based on evidence and considering the different, interrelated factors that may influence learning.

More research is needed to further specify the treatment parameters associated with VAULT. The findings of our early efficacy study have demonstrated the efficacy of the VAULT protocol in both higher and lower dose number per word per session conditions. There are two potential directions for this research to take. One might assume, having established early efficacy, that we would move to a later efficacy stage in which the current VAULT protocol and the specific dose numbers evaluated in this study are examined with a larger number of participants in more generalizable conditions. This is one option. However, there are still many questions regarding optimal treatment provision in terms of specifying session frequency, session duration, and cumulative treatment intensity. The next steps in further evaluating the efficacy of the VAULT intervention will also include early efficacy studies to identify a standard VAULT protocol.

### Limitations

One limitation of the current work is that the current research questions focus solely on the within- and outside-treatment outcome measures without analyzing other “softer” features of the treatment, such as how parents, children, and clinicians perceived the treatment. Another potential limitation of the treatment is that, although the parents were blinded to the treatment targets, many were likely able to ascertain them from context. While we cannot guarantee that this did not affect the way parents filled out the MCDI, the ratio of target/control words produced was quite similar for our directly observed measures (probes: average target 1.54/average control 1.04 = 1.48) and the MCDI reports (average target 3.58/average control 2.54 = 1.40). Thus, it does not appear that parents’ inferred knowledge of target/control words strongly biased their reporting. These are important considerations that can be incorporated into future work. Further research in these areas is warranted as it may be the case

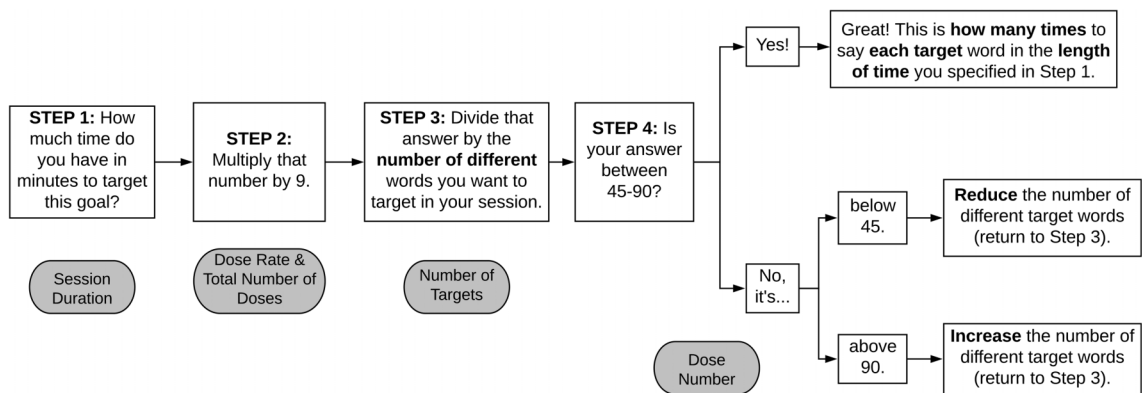
**Figure 6.** A representation of the relationship between dose and several other treatment parameters.

Treatment parameters for a single session	Number of minutes			Number of doses per minute	=	Total Number of Doses	=	Number of therapy targets			Number of doses per target
	Session Duration	×	Dose Rate	Number of Targets				×	Dose Number		
<b>Current study</b>	30 min	×	9 doses per minute	=	270 doses	=	3 targets	×	90 doses per target		
	30 min	×	9 doses per minute	=	270 doses	=	6 targets	×	45 doses per target		
<b>Alternative possibilities</b>	15 min	×	9 doses per minute	=	135 doses	=	3 targets	×	45 doses per target		
	20 min	×	9 doses per minute	=	180 doses	=	4 targets	×	45 doses per target		

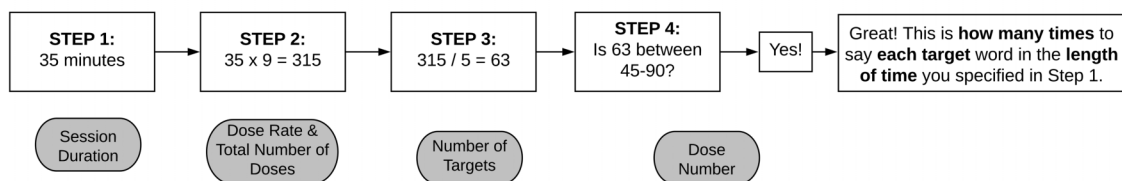
that those other features of treatment affect treatment efficacy as well, which may in turn lead to additional recommendations for clinicians. These types of considerations also point to another area for future exploration: individual differences in response to treatment. While the treatment overall was effective, not every child responded to the treatment. It could be valuable to explore individual child characteristics, word characteristics, or interactions between children and words that might predict response to intervention.

While we would love to be able to report on participant characteristics that indicate which children were likely to be responders or nonresponders or which children seem more suited to the higher or lower dose number condition, the data available reveal no clear patterns relative to age, sex, nonverbal IQ, vocabulary size at treatment initiation, or any other measure we had. Some might wonder about the two participants who had some exposure to another language at home and whether they might have responded differently than the other children in the study. Both children

**Figure 7.** A flowchart to demonstrate how to determine how many words to say how many times.



**EXAMPLE**



were classified as responders. One's effect size was within 1 *SD* of the mean for the treatment condition, while the other had the highest treatment effect size of all the participants. The child with the highest effect size was also a "fast" word learner as measured by the MCDI (and learned more than 10 times as many words per week during and after treatment compared to the delay period prior to treatment). At the same time, the child with the average response rate reached the "fast" rate only at the follow-up appointment, although number of words learned per week at the posttest was three times as fast as during the delay. However, given that these are only two children, it is difficult to generalize these data. While the outcomes for these two participants were positive, we do not know if this is related to their language backgrounds or to some other characteristic.

It is important to remember that we only targeted words the children were reported to understand. Thus, we do not know how VAULT will work for teaching words that are not in a child's receptive vocabulary. This treatment was not intended to address issues with speech sounds, morphology, pragmatics, or other speech and language needs that children may have, and as such, it remains to be seen how VAULT will fit into a more comprehensive treatment program.

## Conclusion

The VAULT treatment protocol was an efficacious way to target expressive vocabulary for late-talking monolingual English-speaking toddlers. As a group, children began producing the targeted words, and more than a third of the children demonstrated evidence of "learning how to learn" words outside the clinic. Clinicians who choose to use this protocol have the freedom to vary the dose number per target word per session between 45 and 90 doses per word. Recall that these dose numbers are based on a consistent dose rate of nine words per minute. Clinicians should also keep in mind that changing the number of words targeted in a session might change the required length of the session. This will allow clinicians to best suit the individual needs of their clients, as there was no difference between outcomes in the higher and lower dose number conditions.

## Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant 1R01 DC015642-01, awarded to Mary Alt and Elena Plante.

We are forever grateful to all the families who trusted us with their beautiful children and participated in this research. We are indebted to the wonderful members of the L4 Lab for all of their contributions from creative crafting, to fidelity checks, to skillful interventions. The work would not be possible without you. We also thank Sarah Cretcher and Isabel Navarro for inventing the Vocabulary Acquisition and Usage for Late Talkers monitor and Elena Plante for her input on this article.

## References

- Alt, M., Meyers, C., 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. <https://doi.org/10.1016/j.jcomdis.2014.07.002>
- American Speech-Language-Hearing Association. (2019). *Late language emergence*. Retrieved from [https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589935380&section=Incidence\\_and\\_Prevalence](https://www.asha.org/PRPSpecificTopic.aspx?folderid=8589935380&section=Incidence_and_Prevalence)
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321–324. <https://doi.org/10.1111/1467-9280.00063>
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). Bloomington, MN: Pearson.
- Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16*, 161–169. <https://doi.org/10.1007/s11065-006-9013-7>
- Buschmann, A., Multhauf, B., Hasselhorn, M., & Pietz, J. (2015). Long-term effects of a parent-based language intervention on language outcomes and working memory for late-talking toddlers. *Journal of Early Intervention, 37*(3), 175–189. <https://doi.org/10.1177/1053815115609384>
- Cable, A. L., & Domsch, C. (2011). Systematic review of the literature on the treatment of children with late language emergence. *International Journal of Language & Communication Disorders, 46*, 138–154. <https://doi.org/10.3109/13682822.2010.487883>
- Campbell, M., Katikireddi, S. V., Hoffmann, T., Armstrong, R., Waters, E., & Craig, P. (2018). TIDieR-PHP: A reporting guideline for population health and policy interventions. *The BMJ, 361*, 1–5. <https://doi.org/10.1136/bmj.k1079>
- Centers for Disease Control and Prevention. (2018). *Important milestones: Your baby by two years*. Retrieved from <https://www.cdc.gov/ncbddd/actearly/milestones/milestones-2yr.html>
- Dale, P. (2007). *MacArthur-Bates Communicative Development Inventories—Third Edition (MCDI-III)*. Baltimore, MD: Brookes.
- Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders, 43*(4), 361–389. <https://doi.org/10.1080/13682820701546854>
- DeVeney, S. L., Cress, C. J., & Reid, R. (2014). Comparison of two word learning techniques and the effect of neighborhood density for late talkers. *Communication Disorders Quarterly, 35*(3), 133–145. <https://doi.org/10.1177/1525740113516788>
- DeVeney, S. L., Hagaman, J. L., & Bjornsen, A. L. (2017). Parent-implemented versus clinician-directed interventions for late-talking toddlers: A systematic review of the literature. *Communication Disorders Quarterly, 39*(1), 293–302. <https://doi.org/10.1177/1525740117705116>
- Ellis Weismer, S., Murray-Branch, J., & Miller, J. F. (1993). Comparison of two methods for promoting productive vocabulary in late talkers. *Journal of Speech and Hearing Research, 36*(5), 1037–1050. <https://doi.org/10.1044/jshr.3605.1037>
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories* (2nd ed.). Baltimore, MD: Brookes.
- Fey, M. E., & Finestack, L. H. (2008). Research and development in child language intervention: A five-phase model. In R. G. Schwartz (Ed.), *Handbook of child language disorders* (pp. 513–529). New York, NY: Psychology Press.



- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*, 677–694. <https://doi.org/10.1017/S0305000916000209>
- Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, *40*, 621–632. <https://doi.org/10.1037/0012-1649.40.4.621>
- Girolametto, L., Pearce, P. S., & Weitzman, E. (1996). Interactive focused stimulation for toddlers with expressive vocabulary delays. *Journal of Speech and Hearing Research*, *39*(6), 1274–1283. <https://doi.org/10.1044/jshr.3906.1274>
- Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation—Second Edition (GFTA-2)*. San Antonio, TX: Pearson.
- Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., . . . Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *The BMJ*, *348*, g1687. <https://doi.org/10.1136/bmj.g1687>
- Individuals with Disabilities Education Improvement Act of 2004 (IDEA), Pub. L. No. 108–446, 118 Stat. 2647 (2004).
- JASP Team. (2018). JASP (Version 0.9.1.0) [Computer software].
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children* (2nd ed.). Bloomington, MN: Pearson.
- Kruschke, J. K. (2013). Bayesian estimation supersedes the *t* test. *Journal of Experimental Psychology: General*, *142*, 573–603. <https://doi.org/10.1037/a0029146>
- Leonard, L. B., Schwartz, R. G., Chapman, K., Rowan, L. E., Prelock, P. A., Terrell, B., . . . Messick, C. (1982). Early lexical acquisition in children with specific language impairment. *Journal of Speech and Hearing Research*, *25*(4), 554–564. <https://doi.org/10.1044/jshr.2504.554>
- MacRoy-Higgins, M., & Montemarano, E. A. (2016). Attention and word learning in toddlers who are late talkers. *Journal of Child Language*, *43*, 1020–1037. <https://doi.org/10.1017/S0305000915000379>
- Plante, E., & Gómez, R. L. (2018). Learning without trying: The clinical relevance of statistical learning. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 710–722. [https://doi.org/10.1044/2018\\_LSHSS-STLT1-17-0131](https://doi.org/10.1044/2018_LSHSS-STLT1-17-0131)
- Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., . . . Burton, R. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology*, *23*, 530–545. [https://doi.org/10.1044/2014\\_AJSLP-13-0038](https://doi.org/10.1044/2014_AJSLP-13-0038)
- 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. [https://doi.org/10.1044/2017\\_LSHSS-17-0077](https://doi.org/10.1044/2017_LSHSS-17-0077)
- Random.org. (n.d.). [Homepage]. <https://www.random.org/>
- Rescorla, L. (2011). Late talkers: Do good predictors of outcomes exist? *Developmental Disabilities Research Reviews*, *17*(2), 141–150. <https://doi.org/10.1002/ddrr.1108>
- Rice, M. L., Buhr, J., & Oetting, J. B. (1992). Specific-language-impaired children's quick incidental learning of words: The effect of a pause. *Journal of Speech and Hearing Research*, *35*(5), 1040–1048. <https://doi.org/10.1044/jshr.3505.1040>
- 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. <https://doi.org/10.1044/jshr.3701.106>
- Robertson, S. B., & Ellis Weismer, S. (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1234–1248. <https://doi.org/10.1044/jslhr.4205.1234>
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374. <https://doi.org/10.1016/j.jmp.2012.08.001>
- Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomized trials. *BMJ Medicine*, *340*, 698–702.
- Singleton, N. C. (2018). Late talkers: Why the wait-and-see approach is outdated. *Pediatric Clinics of North America*, *65*(1), 13–29. <https://doi.org/10.1016/j.pcl.2017.08.018>
- Solomon-Rice, P. L., & Soto, G. (2014). Facilitating vocabulary in toddlers using AAC: A preliminary study comparing focused stimulation and augmented input. *Communication Disorders Quarterly*, *35*(4), 204–215. <https://doi.org/10.1177/1525740114522856>
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*, 58–76. <https://doi.org/10.3758/s13423-017-1323-7>
- Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*, 70–77. <https://doi.org/10.1002/mrdd.20139>
- Wilcox, M. J., Kouri, T. A., & Caswell, S. B. (1991). Early language intervention: A comparison of classroom and individual treatment. *American Journal of Speech-Language Pathology*, *1*(1), 49–62.
- Williams, K. T. (2007). *Expressive Vocabulary Test—Second Edition (EVT-2)*. San Antonio, TX: Pearson.
- Yu, C., Suanda, S. H., & Smith, L. B. (2017). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*, 1–12. <https://doi.org/10.1111/desc.12735>
- Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1562–1592. [https://doi.org/10.1044/1092-4388\(2007/106\)](https://doi.org/10.1044/1092-4388(2007/106))

---

## Appendix A

### Procedures When Children Were Exposed to a Language in Addition to English in the Home

---

Two participants' parents reported that their child spoke a language in addition to English (Haitian Creole, Japanese). After clarifying with the parents that they were comfortable that the treatment would be administered in English only, we proceeded with our assessment and treatment with the following modifications. For these participants, we scored the MCDI and the EVT-2 in the following ways:

1. We first obtained a standard score based on the number of words produced in English only.
  2. We then obtained a total conceptual vocabulary score based on the total number of concepts for which the child produced a word, either in English or in the child's other language. These assessments are not intended to be used for bilingual participants, nor are they designed to be scored in this way; therefore, for the MCDI, we only used the total conceptual vocabulary score to ensure that the child's total vocabulary still fell below the 10th percentile for their age. For the EVT-2, we used this information to better counsel families after the study regarding their child's progress and plans for future therapy outside our program.
  3. During the treatment sessions, we asked the parent to translate any words the child said in another language. These productions were included in our total word counts, although a child never had a target word that was in a language other than English.
- 

## Appendix B

### Reliability Scoring for Participant 10

---

Due to Participant 10's high volume of language produced relative to the other participants, reliability was assessed differently. As usual, a reliability tracker attended every session and transcribed the child's utterances online. However, a second individual observed all videos of the 16 treatment sessions and transcribed the utterances offline. This second individual noted when she agreed with utterances transcribed by the original reliability tracker. Additionally, she noted when she disagreed with the original reliability tracker's transcription, making note of the specific difference, as well as when she heard an utterance that had not previously been heard or recorded. Two new final individuals resolved disagreements between the first two individuals. Their contribution ensured that each of the child's utterances was confirmed by at least two people. These final two individuals each checked reliability for half of the videos. They checked for utterances where there was disagreement between the first and second individuals as well as for utterances that had only previously been heard by the second individual. Whenever the third individual disagreed with both prior transcriptions or disagreed with the new utterances heard by the second person, the other final individual acted as a fourth listener and final decision maker. Ultimately, for each of the child's finalized utterances, there was agreement between at least two individuals.

---