INVESTIGATION OF TREATMENT DOSE SCHEDULE FOR CHILDREN WITH SPECIFIC LANGUAGE IMPAIRMENT

by

Christina N. Meyers

Copyright © Christina Meyers 2015

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF SPEECH, LANGUAGE, AND HEARING SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2015
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Christina N. Meyers, titled Investigation of Treatment Dose Schedule for Children with Specific Language Impairment and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Elena Plante  
Date: March 13, 2015

Mary Alt  
Date: March 13, 2015

Edwin Maas  
Date: March 13, 2015

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Elena Plante
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Christina N. Meyers
ACKNOWLEDGEMENTS

I wish to express my utmost gratitude for my mentors, Dr. Plante and Dr. Alt who have guided and supported me unfailingly for many years during my doctoral studies. Each of you has motivated and encouraged me, seen me through challenges, and celebrated successes with me.

Dr. Plante, I am so very grateful to have you as a mentor. The time you have dedicated to me and what you have taught me far exceeds what I could have ever imagined.

Dr. Alt, you are always willing to go the extra mile. I cannot thank you enough for your enthusiastic support, helping me to develop as a clinician-researcher and guiding me towards so many wonderful learning opportunities.

Dr. Maas, I thank you sincerely for your brilliant insights that contributed to the enhancement of this work.

I appreciate each person who has contributed to the wonderful experience I have had at The University of Arizona—it will always be my home, and will be in my memories.

I also wish to express my greatest appreciation for my family, especially my loving and endlessly supportive parents, Vickie and Chris Meyers. My family has always encouraged me to pursue my interests and dreams, and laid the foundation that has allowed me to do so.

I would like to acknowledge the funding sources that supported me during my doctoral study, including NIDCD -T32DC009398 NIH Grant and the Grunewald-Blintz Foundation Fellowship, and the (Plante) 1R21DC014203 that supported this doctoral dissertation study.
DEDICATION

This work is dedicated to families of children with developmental language disorders. We are working to improve clinical practices to enhance life-long outcomes for all.
# TABLE OF CONTENTS

LIST OF FIGURES ............................................................................................................. 7
LIST OF TABLES .................................................................................................................. 8
ABSTRACT ........................................................................................................................... 9

INTRODUCTION .................................................................................................................. 10
   Specific Language Impairment.......................................................................................... 10
   Treatment Research in Specific Language Impairment.................................................. 11
   General Treatment Parameters...................................................................................... 14
   Massed vs. Spaced Effects on Learning......................................................................... 16
   Conversational Recast.................................................................................................... 22
   Study Purpose.................................................................................................................. 28

METHOD ............................................................................................................................. 30
   Participants ....................................................................................................................... 30
   Procedures ....................................................................................................................... 34
   Experimental Control...................................................................................................... 45

RESULTS ............................................................................................................................. 50
   Generalization Probes .................................................................................................... 53
   In-Treatment Performance ............................................................................................. 55
   Spontaneous Use.............................................................................................................. 56
   Individual Results .......................................................................................................... 57
   Long-term Retention........................................................................................................ 63

DISCUSSION ......................................................................................................................... 65
   Study Findings ................................................................................................................ 65
   Outcome Measures......................................................................................................... 72
   Factors Inherent to the Treatment that May Influence Outcomes................................. 75
   Factors Inherent to the Child that May Influence Outcomes.......................................... 78
   Limitations ....................................................................................................................... 81
   Conclusion ....................................................................................................................... 84

APPENDIX A ......................................................................................................................... 86
REFERENCES ....................................................................................................................... 87
LIST OF FIGURES

FIGURE 1. Average effect size for target and control morphemes by Massed and Spaced treatment condition .................................................................52

FIGURE 2. Spontaneous use of target morpheme during treatment sessions for Massed and Spaced conditions ........................................................................57

FIGURE 3. Single subject data ..........................................................................................................................................................................................59
LIST OF TABLES

TABLE 1. Demographic Information and Standardized Test Scores.............................33

TABLE 2. Treatment Variables for Children in Spaced and Massed Conditions..........37

TABLE 3. Treatment Fidelity: Scores by Treatment Fidelity Components..................49

TABLE 4. Performance data for children in the Spaced and Massed Treatment Conditions ..................................................................................................................................................................52

TABLE 5. In-Treatment Target Morpheme Use & Spontaneous Use for Children in Spaced and Massed Conditions........................................................................................................................................................................58

TABLE 6. Performance: End Treatment and Long-term Retention..............................64
Abstract

Dosage has been identified as important element of intervention that has the potential to affect intervention efficacy. The purpose of this study was to examine the role of dose schedule for treatment of grammatical morphology deficits in children with Specific Language Impairment (SLI). Sixteen 4-5 year old children with SLI participated in a 6-week intervention program during which children received equivalent daily Enhanced Conversational Recast treatment targeting grammatical morpheme errors. Half of the children received treatment in one 30-minute session (massed condition). The other half received treatment in three 10-minutes sessions (spaced condition) over a 3-hour period. Progress was assessed three times weekly by probing a child’s use of his/her treatment morpheme and untreated morpheme (a maturational control) in untreated contexts. Pre-to-post treatment morpheme usage differed significantly for children regardless of dosage condition, demonstrating overall treatment efficacy. There were no differences in treatment effects for the massed and spaced conditions. In addition, nonverbal IQ and receptive vocabulary test scores correlated with treatment effect sizes. The study adds to evidence that Enhanced Conversational Recast can produce positive results, in a relatively short period of time, for children with specific language impairment. Moreover, it appears that clinicians may have some flexibility in terms of the dose schedule they employ to deliver this treatment in an evidence-based manner.
Introduction

Over 50 percent of speech-language pathologists (SLPs) work in school settings and approximately 90% of these SLPs serve children with language impairment (Brandel & Loeb, 2012). For these clinicians, evidence-based information is key to their ability to provide the most effective therapy. The overall cost (in terms of time, money, and other considerations) of intervention to the child, family, providers, and institutions can be reduced if effective and efficient interventions are utilized. Additionally, the potentially negative impact on the child that comes from being removed from the regular education environment frequently, or for an extended period of time, for the purpose of receiving language intervention (e.g., over the course of an entire academic year) can be reduced. Therefore, how to make language intervention more effective, for more children, in less time is of central importance.

Specific Language Impairment

Specific language impairment (SLI) is a language-learning disorder that affects approximately 7-13% of kindergarten age children, making it one of the most common developmental disorders (Tomblin, Records, Buckwalter, Zhang, Smith, & O’Brien, 1997). Children with specific language impairment do not have another diagnosis that would explain their deficits in language development. During the preschool years, SLI is characterized by particular difficulty acquiring and using grammatical morphemes. This includes omission of morphemes that mark both tense and agreement (e.g., present third person singular –s, past tense –ed, and both copula and auxiliary is, are, am, was, and were). Although verb morphology errors have been called a hallmark deficit in preschool age children with SLI (Rice, Wexler, & Cleave, 1995; Leonard,
1998), many children with SLI also have difficulties with pronouns, plurals, possessives, and Wh-question constructions.

Children with specific language impairment can face life-long difficulties including struggles with literacy, academics, and even social and vocational outcomes. For example, Tomblin, Zhang, Buckwalter, and O'Brien (2003) followed a cohort of children with a diagnosis of primary language impairment (an alternate term for SLI) beginning in kindergarten. They found that the poor language skills of children with language impairment were very likely to persist when re-evaluated over time (e.g., Tomblin et al., 2003; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). This suggests that SLI is persistent in nature. Additionally, children with SLI struggle to communicate as effectively as same-age peers, which can have an immediate impact on quality of life. Marton, Abramhoff, and Rosenzweig (2004) reported that 7 to 10 year old children with SLI scored significantly lower than typically developing peers on measures of social cognition. A recent study provided evidence that, even with at least some appropriate support and despite participation in post high school education, young people with a history of SLI have poorer educational attainment and occupational outcomes compared to peers without a history of SLI (Carroll & Dockrell, 2012). The immediate and long-term consequences of SLI indicate that providing effective early intervention is crucial.

Treatment Research in SLI

The American Speech-Language-Hearing Association’s (ASHA) describes the relevance of Evidence Based Practice (EBP) to clinical issues in speech-language pathology and provides guidelines for the principles and procedures involved in EBP
(ASHA, 2004). Of primary importance are data-based reports regarding specific treatment methods that clinicians may wish to adopt for use. While speech-language intervention for SLI is typically effective to some degree, reviews of published treatment studies reveal that treatment outcomes have been mixed at best, and effects are often modest (Law, Garrett, & Nye 2004; Schooling, Venedikov, & Leech, 2010). Positive intervention outcomes are associated with relatively long intervention durations, typically greater than two months to address a single objective (Law et al., 2004). Treatments designed to address expressive morpho-syntax errors appear to be particularly challenging in terms of obtaining clinically useful effects (Law et al., 2004). Reviews of child language treatment literature have led to repeated calls for research designed to produce more effective and efficient interventions (Fey & Finestack, 2009; Law et al., 2004; Proctor-Williams, 2009).

Many researchers (e.g., Robey & Schultz, 1998; Robey, 2004; Pring, 2004) have pointed out that the field of speech-language pathology has not established a systematic framework to develop and evaluate interventions and a significant body of treatment evidence is just not available. Other disciplines (e.g., medicine) have implemented a four or five phase approach to examine outcomes of clinical procedures. The lack of such a framework for behavioral treatment research contributes to its lack of cohesion (Fey & Finestack, 2009).

To address this need, Fey and Finestack (2009) have described a hierarchy of treatment studies that includes pretrial studies, feasibility studies, early and later efficacy studies, and effectiveness studies. Feasibility studies are designed to be a first test of a specific hypothesis regarding intervention strategies. Their purpose is to evaluate the clinical applicability and viability of a particular intervention technique, as
well as evaluating the outcome of its use. Less importance is placed on the magnitude of the outcome during a feasibility study, but positive effects are generally expected because well-motivated feasibility studies are based on clinical experience and evaluation of evidence gathered in previous research.

Efficacy studies examine outcomes of a treatment technique under ideal conditions. *Early efficacy studies* are designed to establish a causal relationship between an intervention technique and an outcome under well-controlled conditions. The important features of an early efficacy study are internal validity and replicability. Early efficacy studies may be experimental, or quasi-experimental in which participants are not randomly assigned to conditions. *Later efficacy studies* are also designed to establish a causal relationship between a treatment variable and an outcome, but do so under more generalizable conditions. Later efficacy studies may also examine a combination of intervention components that have previously been established as efficacious in an early efficacy study. Additionally, later efficacy studies may compare a set of intervention components to another established treatment technique (the ‘standard’). The purpose of both early and later efficacy studies is to demonstrate a relationship between an intervention and an outcome under rigorously controlled experimental conditions (Fey & Finestack, 2009).

In contrast to efficacy studies, *effectiveness studies* can be conducted once a treatment has been demonstrated to be efficacious under ideal conditions (Dollaghan, 2007; Robey & Schultz, 1998). Treatment effectiveness is then evaluated in terms of real-world implementation. Despite its importance for evidence-based clinical decision-making, there is very little treatment effectiveness information available to date (e.g., Dollaghan, 2007). This is due in part to the small number of efficacy studies
that have produced effects that justify translation to the effectiveness research stage (Law et al., 2004; Schooling, Venedikov & Leech, 2010). In particular, treatment studies targeting morphosyntax, a hallmark deficit for preschool children, are very few in number (Schooling et al., 2010).

General Treatment Parameters

Many factors contribute to the cohesiveness of the body of language intervention research (Fey & Finestack, 2009). Treatment studies are complex and multi-faceted by nature. Treatment studies can target any of a wide variety of goals and the children treated can be heterogeneous in relation to these goals. These factors are inherent to treatment research. However, the literature also suffers from limited reporting of the details necessary to evaluate or replicate treatments. In particular, factors related to specification of the treatment dose, the treatment intensity, and the schedule over which treatment is administered are often underspecified.

Treatment dose. Warren, Fey, and Yoder (2007) published a detailed framework for the examination of components of dosage, including definitions of treatment dose, dose frequency, and measures of intensity (discussed below) in an effort to facilitate discussion of the importance of specifying these features in intervention study design and reporting. They define dose as the number properly administered treatment units. The treatment unit, the dose form, is the set of actions that are believed to effect change in the child. In the case of Conversational Recast treatment, elements of the recast procedure are the dose form and each recast would constitute one dose.
**Treatment Intensity.** Warren et al. (2007) describe several parameters that together define treatment intensity. *Dose frequency* is the number of times per week the treatment occurs. The *total intervention duration* is the total time period of the therapy program. The *cumulative intervention intensity* is the product of the dose by the dose frequency by the total intervention duration. Warren et al. noted that previous research has not provided clear definitions or adequate consideration of these treatment parameters to enable evaluation of the effects of intensity.

As exemplified by the title of Yoder, Fey, and Warren's 2012 commentary "Studying the impact of intensity is important but complicated", it is widely recognized that the effect of intensity on treatment outcomes should be carefully explored because intensity manipulations could influence intervention efficacy and efficiency. However, it is difficult to isolate individual factors that contribute to the multifaceted concept of intensity, and difficult to make conclusions about factors across different treatment methods, treatment goals, and clinical populations.

**Treatment Procedures and Goals.** Baker (2012) further considered the issue of optimal intensity for interventions. One practical issue identified by Baker that complicates the study of intervention is *what* and *how* goals are targeted: what is the unit of delivery of therapeutic input (e.g., conversational recast, clinician models), or actions of the client (e.g., naming, imitation) that are being measured? These are inherent to the concept of *dose form*. Given considerable variety across subfields of speech-language pathology, and the myriad of components that could be measured, it is clear that additional operational definitions will be necessary to make comparisons of intensity across intervention studies. Yoder, Fey, and Warren (2012) also recognized that there are various ways that dosage effects on speech-language
development could be studied. They questioned if perhaps a narrow focus on a specific aspect of intensity (such as dose frequency) is inadequate and may ultimately only uncover an effect that is minimal relative to the overall outcome. They suggested that an additional variable of treatment intensity may be the “spacing of teaching episodes within an intervention session.” They conclude by stating that spacing may be a highly influential aspect of treatment intensity (Yoder et al., 2012). This indicates the need to manipulate dose delivery schedule, or the spacing of doses, independent of dosage or cumulative intensity, to describe the effect of dose schedule on treatment outcomes.

Massed vs. Spaced Effects on Learning

One common distinction between dose delivery schedules is whether doses are delivered clustered within a compressed period of time or whether they are spaced across a longer period of time. Effects of distributed and massed practice for many types of skills have been studied for over 100 years in numerous contexts and across various domains of perceptual learning (e.g., visual, verbal, and motor tasks). In the linguistic domain, a meta-analysis of 93 studies demonstrated the benefit of distributed learning for a wide variety of linguistic targets, including real or novel words and sentences, delivered in various modalities (e.g., auditory, visual) to adults and children (Janiszeweski, Noel, & Sawyer, 2003). However, none of the studies included in this review included children with communication disorders. Maas and colleagues reviewed motor learning literature and presented principles that could potentially be translated to clinical treatment of apraxia of speech (Maas, Robin, Austermann Hula et al., 2008). These authors also highlighted the multi-faceted nature of this research by defining five conditions of practice: amount, distribution, variability,
schedule, and attentional focus. The relative effect of dose delivery schedule may ultimately depend on these variables as well as what is being treated, and the potentially an interaction of these complex variables in combination.

Many studies of massed vs. spaced learning have involved ‘verbal learning’, in which material is presented and tested verbally. This usually involved learning of paired associates or word lists. Often these studies measure recall for exact items, thus testing rote memory. Current literature on massed and spaced learning generally indicates that there is an effect of enhanced learning that occurs following spaced practice (Janiszeweski et al., 2003). This is likely very different from how language is learned either naturally or in the context of language therapy. Therefore the relevance of these studies to the more naturalistic contexts is unclear. Furthermore, learning words is different from learning morphosyntax because instead of involving repetitions of stable phonological forms, learners are exposed to various exemplars of grammar and must extract a rule from the given exemplars.

A few studies have examined the effect of spacing on learning when information must be abstracted from multiple exemplars. This is a context that is more similar to that involved in learning grammatical forms. Vlach, Sandhofer, and Kornell (2008) used a category induction task with 3-year-old children. Children were presented with pictures of objects that were exemplars of discrete categories. Children performed better on a follow-up test for category exemplars that had been presented in a spaced manner than exemplars that had been presented in a massed sequence. The benefit of spaced learning trials occurred even though spacing may have increased the difficulty of the task by allowing children time to forget previous exemplars (Vlach et al., 2008).
In another study, exemplar spacing was found to enhance retention and transfer. Three to six month old infants were presented exemplars in either a spaced manner—across multiple days, or massed manner—all in one day (Merriman, Rovee-Collier, & Wilk, 1997). As in the Vlach et al. study, these exemplars represented different categories of items. The infants’ ability to recognize the same exemplar they had seen a day earlier, considered retention, was enhanced in the spaced condition relative to the massed condition. Three month olds did not remember the exemplar the next day if they had been given only a massed exposure, which was in agreement with well-established general memory literature (e.g., Crowder, 1976). In contrast, six-month-old infants were able to recognize a novel category member after 24 hours. In an additional experiment reported by Merriman et al., infants at both ages who were shown exemplars of a category in a spaced manner were able to generalize to a novel category member, which was considered transfer. Overall, the authors concluded that greater spacing between exemplars promoted retention.

Studies of Intervention Schedule in Language Treatment Contexts. Ukrainetz, Ross, and Harm (2009) tested the effects of two intervention schedules for pre-literacy training for at-risk kindergartners in a school setting. Forty-one 5-6 year old children with low phonemic awareness were provided training for 11 hours, either massed into three times per week, and spaced at one time per week over a longer period of time. The effects were measured after the training period, and at a follow up months later. There were no differences in the outcome for the two dose schedules at either of the time points. Finally, by the end of the school year, there were no differences in phonemic awareness between either the massed, spaced, or a no-treatment control condition. The long-term retention results suggest that the learning that occurred may
have been tenuous and eroded with time. If so, it is difficult to draw any conclusions about the relative effects of treatment schedule from this study.

Some existing treatment studies (e.g., Gray 2003; Plante et al., 2014) demonstrate that enhanced learning for children with SLI can occur following massed practice, but have not considered the effect of spaced input. Children with frank language learning impairments need more exposure to new linguistic forms to be learned than do typically developing children (e.g., Oetting, Rice & Swank, 1995). High-density presentation, which provides more exposures per unit of time, is easier to provide in massed than spaced dose schedules. However, Riches, Tomasello, and Conti-Ramsden (2005) suggested that both high-density input and more spaced input are needed for learning and retention. The first may help to establish initial representations, and the second may reinforce these representations. Thus, there may be a role for both types of input, and perhaps, at different points in time.

Smith-Lock and colleagues (Smith-Lock, Leitão, Lambert, Prior, Dunn, Cronje et al. 2013a) examined the effect of dose frequency while keeping cumulative intervention intensity (in minutes) constant. They implemented a grammar treatment program they had previously developed for use with 5-year-old children with SLI (Smith-Lock, Leitão, Lambert, & Nickel, 2013b). The basic treatment consisted of direct instruction, imitation, focused stimulation, and recasting. This combination of treatment procedures produced a significant treatment effect over 8 weeks with a large overall average effect size of $d=1.24$, or an average of $d=1.66$ for children without articulatory limitations (Smith-Lock et al., 2013b). In their subsequent study (Smith-Lock et al., 2013a), dose frequency was manipulated to create two conditions: a massed condition, in which treatment was administered in one-hour sessions one time
per day for 8 days, and a spaced condition, in which one-hour sessions were provided one time per week for eight weeks. Results showed that children who received treatment weekly outperformed children who received treatment daily. Although the study was designed to avoid confounds of different dose forms and cumulative intervention intensity in minutes, there are caveats to the type of design they used. This study can be conceptualized as providing massed intervention (eight consecutive days) or spaced intervention (eight days distributed over eight weeks). However, this design resulted in two very different treatment durations (just over a week versus eight weeks). Additionally, the authors stated that dosage—the number of teaching episodes was not controlled and may have been discrepant between groups (Smith-Lock et al., 2013a).

In the domain of lexical learning, Riches, Tomasello, and Conti-Ramsden (2005) presented novel verbs to children with SLI through either 12 presentations massed in one session, with 18 presentations in one session, or 12 presentations spaced across four sessions. The children were tested by being asked to demonstrate the action of the novel verb with an object, produce the novel verb, and receptively identify the novel verb. Children learned better in the more distributed (four sessions) condition, even though one of the massed conditions actually provided more presentations (i.e., a higher dose). However, this effect was only significant for children’s production of new words, not for their recognition of them. This production difference, however, is consistent with the statement by Childers and Tomasello (2002), that “given an equal number of exposures, distributed (or spaced) practice at skills is almost always superior to massed practice with skills” (p.968).
Barratt, Littlejohns, and Thompson (1992) conversely demonstrated that a more intensive treatment schedule (massed treatment) led to greater improvements in expressive language outcomes compared to less intensive intervention (spaced treatment). Their two- to five-year-old participants received therapy either four times weekly in a three week block within each of two three-month cycles, or one time weekly for six months. In this way, overall treatment duration was partially controlled at approximately 6 months. The dose number and cumulative intervention intensity were held constant. However, the dose frequency and related spacing of treatment sessions was unequal between conditions. Although this study lacked strong experimental control over these parameters, the authors concluded that higher intensity treatment led to greater gains that the lower intensity treatment (Barratt et al., 1992).

In contrast, Proctor-Williams and Fey (2007) did not find this same relationship between higher intensity treatment and better outcomes in a study of 7 to 8 year old children with SLI and typical ‘language similar’ children age 5-6 years. They presented children with novel irregular verbs at three recast densities: none, conversational, and intervention, over five intervention sessions. This study showed that children with SLI did not benefit from recasts presented at a low (“natural”) rate, although children with typical language could. However, when the rate of recasts and models was increased, neither group benefitted and the increased rate had actually had a negative effect on the typical learners. The authors acknowledge that the spacing of the treatments may have contributed to the outcome; they were not able to tightly control the distribution of the stimuli. However, because neither level of recast provided was effective for children with SLI, this training method may have been generally ineffective.
Additionally, because both recasts and models were provided, the effects of these two individual treatment techniques were confounded. Therefore, conclusions that may be drawn about schedule distribution from this study are extremely limited.

Some of the studies of massed and spaced learning make a distinction between immediate performance effects, and long-term retention. Kamhi (2014) has made a similar distinction in terms of treatment effects. Traditionally there has been a distinction between learning and generalization that can be conceptualized as a distinction between immediate performance effects and ‘long term, context-independent’ learning (Bjork, 2004). Additionally, learning may be evaluated in terms of transfer or generalization. Transfer, referred to from here on as generalization, is the ability to apply newly learned skills to novel situations or uses. Kamhi argues that the current conceptualization of generalization may be inaccurately characterizing children’s learning problems as being specific to generalization. Rather, grammatical learning by his definition should imply that the child is spontaneously producing grammatical utterances across contexts and with new (or untrained) vocabulary.

Conversational Recast Treatment

One intervention method that has consistently shown positive outcomes is Conversational Recast treatment. Conversational recast has been a frequently used method with preschool-age children to teach grammatical morpheme use (e.g., Camarata & Nelson, 1992; Camarata, Nelson, & Camarata, 1994; Nelson, Camarata, Welsh, Butkovsky, & Camarata, 1996; Leonard, Camarata, Brown, & Camarata, 2004; Leonard, Camarata, Pawlowska, Brown & Camarata, 2006). Conversational recasts occur in a naturalistic context, during play activities in which the child and clinician are
both engaged. A recast is defined as a clinician utterance that immediately follows a child’s utterance and provides a correct model of that utterance. For example, the child may make statements or ask questions in which a morpheme targeted for remediation should have been used, but was not. The clinician then uses the child’s own words, correcting the grammatical error. The clinician’s response does not add semantic information or necessarily expand the length of the utterance, beyond the inclusion of necessary morphological markers. Recasts can follow child utterances that are either grammatically correct (a confirming recast) or incorrect (a corrective recast). A study by Hassink and Leonard (2010) found that conversational recasts that occurred following a child’s correct or incorrect utterances were facilitative and associated with gains on the target form.

There are several sources of evidence to support the basic efficacy of conversational recast treatment targeting morphosyntax. Conversational recast has been contrasted with imitation treatment, and found to produce superior results. In a study by Camarata, Nelson and Camarata (1994), 4-and 5 year-old children were provided with either conversational recast treatment or imitative treatment. The conversational recast group required significantly fewer presentations of the target (embedded in conversational recasts) to produce the target *spontaneously* and children in the conversational recast condition produced their morpheme targets spontaneously more often than children in other group. In contrast, the imitative treatment group produced more elicited utterances that followed models provided within treatment sessions. In terms of generalization of learning, spontaneous use may be a better measure of learning than the elicited production these children demonstrated. However, as Kahmi (2014) has noted, performance within treatment
sessions is a poor index of true learning. He would consider the elicited productions under the imitative training condition to be an immediate performance effect rather than a demonstration of learning that is likely to be applied outside of the treatment context, explaining low spontaneous use under the imitation condition.

Positive results from Conversational Recast treatment requires an adequate number of recasts, and likely requires the recasts to be distributed over a minimum period of time. Successful experimental recast treatments have spanned 5 weeks (Plante et al., 2014) to six months (Leonard et al. 2006). The treatment effect sizes for these two studies were comparable, though Leonard and colleagues’ treatment was needed for a much longer period of time to induce positive results. As highlighted in Law et al. (2004), even modest treatment outcomes are typically associated with longer (> 2 months) treatment durations.

Likewise, conversational recast treatment delivered over a very short period of time may not be efficacious. Proctor-Williams and Fey (2007) utilized conversational recasts and models of varying intensities, but the children with SLI in their study did not learn from this combination of conversational recast and models. This is a counter-example of the efficacy of conversational recast that is likely due to an insufficient number of target exposures over too short of a period of time (7 exposures in each of 5 sessions).

Treatment studies that require very long intervention periods to produce results can introduce a number of potential confounds. For example, for ethical reasons, children cannot be removed from outside treatments while they participate in treatment research. This can potentially lead to unexpected gains in untreated morphemes. For example, Leonard et al. (2006) reported gains in an untreated
morpheme (3rd person agreement) when children were trained on auxiliary verbs over the course of six months. This could constitute cross-morpheme generalization. However, some of this generalization might not have been treatment induced. Leonard et al. (2006) reported that their subjects were receiving outside therapy, which could account for improvement in morphemes that were not treated within the intervention study.

The long treatment durations in the Leonard et al. (2006) and other studies (e.g., Barratt, Littlejohns, & Thompson, 1992) also raise the possibility that maturational factors may play a role in what appear to be intervention effects. Leonard et al. 2006 indicated that as many as 5 of the children in their study may have been “late bloomers” rather than having a true specific language impairment. If so, their data may have inflated the post-treatment scores. Nevertheless, the authors state that the overall data conformed to the profile seen for the larger group, with the greatest gains seen for the target form and lesser gains on the control form. However, children in a general language stimulation control condition also showed significant gains from Time 1 to Time 4, which the authors note may have been maturation related (Leonard et al., 2006).

Conversational recast treatment is thought to be effective because it occurs in naturalistic contexts and provides an adult model that a child can use to more or less evaluate his or her own utterance (Camarata et al., 1994). However, it is not really known what parameters of treatment are truly the ‘active ingredients’ affecting change in language use. Potential active ingredients include the naturalistic setting, the social interactions involved, the grammatical models provided by the clinician, the spacing of models over time within the conversation, and the opportunity for the child to
compare his own productions to the clinician models. There are other, unknown possibilities as well. Typically, there is a problem of underspecification of treatment methods reported in the literature. For example, although treatment frequency is often reported, few studies report information on the number of treatment doses administered and in which dose form (e.g., the number of recasts), the duration of treatment, and cumulative intervention intensity (Warren et al., 2007; Proctor-Williams, 2009).

Conversational Recast is a promising treatment that may be improved to further increase the efficacy or efficiency of the procedure. One potential avenue for improving effects in conversational recast treatment is to incorporate principles drawn from more general learning theory (Alt, Meyers, & Ancharski, 2012).

Surprisingly, few treatment studies, conversational recast or otherwise, explicitly incorporate information from research on learning, although ample evidence now suggests that it should be. Some researchers in the fields of speech-language pathology have recently focused on intervention components that are known to facilitate rapid learning in typical learners and suggested that these components be translated to therapeutic contexts (Alt et al., 2012; Plante et al., 2014). For example, a recent vocabulary intervention for toddlers with delayed expressive language—“late talkers” demonstrated that the inclusion of the components that benefit learning in typical individuals, such as a high dose of input and enhanced linguistic variability of input, have also led to better intervention outcomes compared to those previously reported in the literature (Alt, Meyers, Oglivie, Arizmendi, & Nichols, 2014).

One principle derived from studies of language learning in cognitive science that could be feasibly incorporated into a wide range of language intervention methods
is that of exemplar variability (Alt et al., 2012; Plante et al., 2014). Artificial language learning studies of exemplar variability have repeatedly shown that high variability input enhances rapid learning (Gómez, 2002, Gómez & Maye, 2007), even for learners who have language impairment (Grunow, Spaulding, Gómez, & Plante, 2006; Torkildsen, Dailey, Aguilar, Gómez, & Plante, 2013). For example, when training verb tense markers, such as the auxiliary “is helping” many verbs would be presented in the “is verb- ing” construction. Thus, the “is-verbing” form is presented using high lexical variability. Plante et al. 2014 was the first known translation of this variability principle to a therapeutic context. They showed that increasing the number of different verbs heard when training the use of grammatical morphemes enhanced learning compared with repeating of a smaller set of verbs within each treatment session. The incorporation of high variability in treatment resulted in effect sizes after less than six weeks of treatment that were comparable to a very similar study obtained after six months of treatment (Leonard et al., 2006).

The treatment methods used in Plante et al. (2014) are now referred to as Enhanced Conversational Recast. In Enhanced Conversational Recast, high variability of variant components of target structures is incorporated to increase the saliency of invariant components of a structure. Additionally, clinicians are directed to use a variety of activities during sessions, to not repeat activities frequently, and are free to use many linguistic forms that are not limited to the target embedded in the conversational recast. This incorporates variability into the materials and training context (cf. Perry, Samuleson, Malloy, & Schiffer, 2010) and facilitates high lexical diversity not only within but also across treatment sessions (Plante et al., 2014). These procedures result in high variability, both in activities and contexts, and in the
language used during treatment sessions. Furthermore, clinicians are encouraged to use extra-linguistic cues, such as gesture and gaze, and may position themselves close to the treated child, usually in his or her field of view, when recasting an utterance. Clinicians use attentional cues (e.g., a light touch, directing the child’s gaze to the clinician’s face, the command “look”) to increase the likelihood that the child will be attending to the recast (i.e., looking at the clinician) when it is administered.

Study Purpose

The purpose of this study was to examine the effects of intervention dose schedule on treatment outcomes for children with specific language impairment. Specifically, the study examined massed versus spaced intervention delivery on an input-based treatment for grammatical morpheme remediation. Under the massed treatment schedule, treatment was delivered during one 30-minute session. Under the spaced treatment schedule, the same treatment was spread across three 10-minute sessions over a few hours.

Given the inconsistent results of both experimental and treatment studies that have contrasted massed vs. spaced learning, it is possible that either condition could lead to superior treatment effects. As discussed above, rapid learning has been demonstrated repeatedly under conditions that include massed presentation of exemplars reflecting grammatical forms (e.g., Gómez 2002, Gómez & Maye, 2007). The Enhanced Conversational Recast study (Plante et al., 2014) also demonstrated that 24 unique recast provided within each intervention session benefitted learning (Plante, et al., 2014). However, that same benefit was not obtained if the massed recasts did not contain the variability essential to promote learning. Furthermore, these children heard a large number of unique lexical items embedded in the conversational recasts
given over the entire course of the study. However, this variability alone did not appear to affect learning as much as the degree of lexical variability that occurred within each treatment session. This finding lends support the notion that massed treatment benefits are possible when the basic treatment is itself effective.

The counter-argument in favor of spaced treatments can also be made. Researchers have attributed the benefit of spaced learning to factors such as greater variability of contextual cues during encoding (Glenberg, 1979) and the greater challenge for memory retrieval when items to be learned are more separated in time (Schmidt & Bjork, 1992). Variability of encoding occurs naturally in spaced trials. If input is provided with some time elapsing between exemplars, each trial will naturally occur in somewhat different contexts (e.g., changes in the environment, different objects or actions co-occurring with each presentation). These contextual differences could support both encoded and provide additional cues that support retrieval. In a simple sense, the longer the exposure period, the more likely that variability will be introduced into the learning context. Analogously, the spacing in time between exemplars may make it more likely to forget the previous one before the next is encountered. Rather than being detrimental, the requirement to recall the previous exemplar is more difficult and requires more practice to do, thus strengthening the memory (e.g., Vlach et al., 2008).

Given that there are plausible reasons to hypothesize that either massed or spaced treatment could result in superior outcomes, the hypothesis under study is necessarily bi-directional. Accordingly, the hypothesis is that massed or spaced dose schedules will differentially affect the efficacy of Enhanced Conversational Recast treatment.
Method

Participants

Sixteen children ages 4;10 (years; months) to 5;10 ($M = 5;3$) with SLI participated in this study. Children (n = 8 per group) were assigned to *massed* or *spaced* delivery conditions. All children participated for the full duration of the intervention program. Children met the standard inclusionary criteria for SLI: language impairment in the absence of other handicapping conditions. Language impairment was indicated by a score of 87 or less on the SPELT-P2 (Dawson, Stout, Eyer, Tattersall, Fonkalsrud, & Croley, 2003), an empirically-derived cut-off score validated for monolingual children (Greenslade, Plante, & Vance, 2006). These scores and scores from additional standardized tests are reported in Table 1.

Participants were primary English speakers, however two children (202 & 502) had considerable exposure to Spanish. In each case, one parent used Spanish predominantly to communicate with the child. However, the parents of both children reported that the child’s primary language was English and that the child had been exposed to English from birth. A licensed and certified bilingual speech-language pathologist, who was not otherwise involved in this study, confirmed language impairment for both of these children. Furthermore, all children had attended a preschool or daycare where English was spoken with the exception of one monolingual English-speaking child (301) who had never been enrolled in a preschool or daycare program.

To rule out the presence of hearing loss, all children passed a pure-tone hearing screening at 20dB at 1000, 2000, & 4000Hz, and 25dB at 500Hz. Intellectual disability was excluded by requiring a non-verbal IQ score of greater than 75 (70+5 SEM) on the
nonverbal scales of the Kaufman Assessment Battery for Children—Second Edition, (K-ABC). Additional exclusionary criteria included parental report of a handicapping condition that could account for poor language (e.g., neurological disorders, hearing loss, autism). A diagnosis of attention deficit hyperactivity disorder (ADHD) was not an exclusionary criterion for the study, given the frequent co-morbidity between SLI and ADHD. However, no child in this study had a diagnosis of ADHD. As descriptive measures of receptive language skills, the Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4) and Test of Language Development—Preschool, 2nd Edition, Grammatical Understanding subtest (TOLD–GU) were also administered to children prior to the start of the program.

Although no child was receiving language therapy at the time of study, 11 of the 16 children were receiving concurrent articulation therapy during the language treatment period. Articulation treatment was provided through the preschool program but did not address the speech sounds specific to the child’s target or control morphemes, nor was grammar corrected or targeted in any way during the articulation therapy.

A higher rate of speech sound errors may be reasonably expected in groups of children with SLI relative to typically developing peers, so children involved in this study needed only to have sufficient speech sound production skills to: 1) produce the target and control morphemes in the appropriate word position(s) and 2) be reasonably intelligible and consistent in speech production such that recasting of the child’s utterance by an adult is possible in most instances. To describe children’s speech sound inventory and to make reasonable choices for treatment targets and controls, children were administered the Goldman-Fristoe Test of Articulation—2nd
Edition (GFTA-2) prior to the start of the treatment program. GFTA-2 scores for children in the spaced condition, did not differ significantly from those in the massed condition, $t(14) = .849, p = .410, d = .362$.

Demographic variables and test scores for children assigned to spaced and massed conditions were tested for group differences using two-tailed t-tests. Neither of the demographic factors (gender, age) differed significantly between groups. The spaced condition consisted of four males and four females; the massed condition consisted of five males and three females. Age did not differ between groups, $t(14) = .742, p = .470, d = .320$. There were no group differences in scores obtained on any standardized test ($p > .05$ for all): SPELT-P2: $t(14) = .0982, p = .934, d = .372$; K-ABC $t(14) = .076, p = .940, d = .037$; PPVT-4: $t(14) = 1.796, p = .0942, d = .866$; TOLD-GU $t(14) = .888, p = .390, d = .343$. No corrections for multiple comparisons were used.
Table 1. Demographic Information and Standardized Test Scores

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Gender</th>
<th>K-ABC</th>
<th>SPELT-P2</th>
<th>PPVT-4</th>
<th>TOLD-GU</th>
<th>GFTA-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>402</td>
<td>4;10</td>
<td>F</td>
<td>132</td>
<td>75</td>
<td>129</td>
<td>15</td>
<td>94</td>
</tr>
<tr>
<td>201</td>
<td>5;4</td>
<td>M</td>
<td>103</td>
<td>78</td>
<td>100</td>
<td>11</td>
<td>91</td>
</tr>
<tr>
<td>202</td>
<td>5;3</td>
<td>M</td>
<td>111</td>
<td>82</td>
<td>110</td>
<td>8</td>
<td>106</td>
</tr>
<tr>
<td>301</td>
<td>5;0</td>
<td>M</td>
<td>104</td>
<td>82</td>
<td>97</td>
<td>7</td>
<td>54</td>
</tr>
<tr>
<td>401</td>
<td>5;10</td>
<td>F</td>
<td>99</td>
<td>61</td>
<td>88</td>
<td>8</td>
<td>50</td>
</tr>
<tr>
<td>302</td>
<td>5;8</td>
<td>M</td>
<td>90</td>
<td>70</td>
<td>104</td>
<td>9</td>
<td>93</td>
</tr>
<tr>
<td>601</td>
<td>5;3</td>
<td>F</td>
<td>105</td>
<td>64</td>
<td>97</td>
<td>9</td>
<td>99</td>
</tr>
<tr>
<td>501</td>
<td>5;5</td>
<td>M</td>
<td>94</td>
<td>73</td>
<td>104</td>
<td>11</td>
<td>45</td>
</tr>
<tr>
<td>Mean</td>
<td>5;4</td>
<td></td>
<td>104.75</td>
<td>73.13</td>
<td>103.63</td>
<td>9.75</td>
<td>79</td>
</tr>
<tr>
<td>SD</td>
<td>3.9 mos.</td>
<td></td>
<td>12.83</td>
<td>7.79</td>
<td>12.13</td>
<td>2.55</td>
<td>24.83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Gender</th>
<th>K-ABC</th>
<th>SPELT-P2</th>
<th>PPVT-4</th>
<th>TOLD-GU</th>
<th>GFTA-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>203</td>
<td>4;11</td>
<td>F</td>
<td>102</td>
<td>79</td>
<td>117</td>
<td>8</td>
<td>50</td>
</tr>
<tr>
<td>304</td>
<td>5;6</td>
<td>F</td>
<td>122</td>
<td>84</td>
<td>100</td>
<td>11</td>
<td>84</td>
</tr>
<tr>
<td>502</td>
<td>5;4</td>
<td>M</td>
<td>107</td>
<td>42</td>
<td>86</td>
<td>10</td>
<td>66</td>
</tr>
<tr>
<td>204</td>
<td>5;6</td>
<td>F</td>
<td>99</td>
<td>84</td>
<td>90</td>
<td>8</td>
<td>53</td>
</tr>
<tr>
<td>403</td>
<td>5;1</td>
<td>M</td>
<td>122</td>
<td>58</td>
<td>93</td>
<td>8</td>
<td>78</td>
</tr>
<tr>
<td>404</td>
<td>5;1</td>
<td>M</td>
<td>107</td>
<td>84</td>
<td>93</td>
<td>9</td>
<td>79</td>
</tr>
<tr>
<td>602</td>
<td>5;0</td>
<td>M</td>
<td>91</td>
<td>60</td>
<td>84</td>
<td>9</td>
<td>96</td>
</tr>
<tr>
<td>305</td>
<td>5;4</td>
<td>F</td>
<td>84</td>
<td>34</td>
<td>82</td>
<td>8</td>
<td>54</td>
</tr>
<tr>
<td>Mean</td>
<td>5;3</td>
<td></td>
<td>104.25</td>
<td>65.63</td>
<td>93.13</td>
<td>8.88</td>
<td>70</td>
</tr>
<tr>
<td>SD</td>
<td>2.7 mos.</td>
<td></td>
<td>13.44</td>
<td>20.14</td>
<td>11.24</td>
<td>1.13</td>
<td>16.81</td>
</tr>
</tbody>
</table>

Note: Age (years; months). F = female; M = male. K-ABC = Kaufman Assessment Battery for Children, 2nd Edition, Nonverbal scale; SPELT-P2 = Structured Photographic Expressive Language Test—Preschool II; PPVT-4 = Peabody Picture Vocabulary Test—4th Ed.; TOLD-GU = Test of Language Development-Preschool, 2nd Ed., Grammatical subtest; GFTA-2 = Goldman Fristoe Test of Articulation-2nd Ed. Test scores are standard scores with a mean of 100 (SD = 15) with the exception of the TOLD-GU (M = 10, SD = 3).
Procedures

*Pretreatment Probe Sessions.* In order to determine appropriate target and control morphemes for each child, and to establish baseline use for these morphemes, the first three consecutive days of the intervention program exclusively involved probes of morpheme use. Potential grammatical morphemes to be probed were identified for each child based on errors the child had previously made on the SPELT-P2, or during a language sample obtained prior to the start of the study. Four to six potential morpheme targets were identified for each child. These morphemes were probed on each of the three pretreatment probe days. During the pretreatment probe sessions, clinicians worked with children in one-on-one sessions. Clinicians chose at least three activities and created multiple contexts in which a child could be expected to use each of the morphemes targeted for probing.

Probes for potential morpheme targets occurred in the context of play activities in which use of a variety of morphemes might be used. Clinicians noted utterances by the child in which targeted morphemes were either used correctly, were omitted, or were used in error (e.g., “him” for “she”, omission of verb tense markers). Clinicians noted any spontaneous attempts by the child to use each morpheme, and whether each was correctly or incorrectly used. For the purpose of this study, *spontaneous use* was defined as correct use of a morpheme that did not immediately follow a clinician’s prompt to produce an utterance, and did not contain the same verb (or lexical item associated with the target) provided in the clinician’s previous utterance.

Some children did not produce many spontaneous attempts to use the targeted morphemes on their own. To increase the number of child attempts to use the targeted morphemes, clinicians also elicited use of targeted morphemes. To do this,
clinicians used linguistic contexts that would normally obligate the child to use a particular morpheme in response to a clinician request for information. For example, to probe for the use of she in a child’s speech, the child could be asked to talk about a female character, person, or toy. (See Appendix A for more examples of obligated contexts). Clinicians were asked to record a minimum of ten data points representing spontaneous or elicited utterances for each morpheme, and whether the child’s utterance was grammatically correct. A data collection form was provided for this purpose. All sessions were audio and video recorded to allow for clinicians to resolve any questions concerning the child’s morpheme use and the clinician’s data form.

Two morphemes were selected among the probed morphemes for use in the experimental protocol. Morphemes selected were those that the child demonstrated he or she was not using or was using at very low accuracy levels (<33%). One morpheme was selected for treatment and the other was tracked, but not treated. Improvement in the use the treated morpheme but not the untreated morpheme would be consistent with an effect of treatment, rather than maturation. The morpheme targeted for remediation (the target morpheme) and the untreated morpheme (the control morpheme) were selected from among the morphemes probed during the first three days. The criteria for selecting morphemes as either target and control forms included a) low accuracy use for three days and b) no increasing trend in accuracy across days. Low accuracy use was defined as 0%-40% accuracy on any pretreatment probe day, with an average of less than 33% combined correct use across the three consecutive probe days. Only one child (404) had pretreatment target use at the high end of this range. All other participants demonstrated 20% or less accuracy in pretreatment target use ($M = 7.5\%$). Control morpheme use during
baseline probes for all children averaged 0%-33% ($M = 15\%$). This morpheme selection procedure resulted in each child having a low and stable baseline for the target and control morphemes in the pre-treatment phase. There were no differences in pretreatment use of the treatment morphemes for children in spaced and massed conditions, $t = .906, p = .380, d = .376$.

Selection of target morphemes prioritized morphemes with the lowest accuracy use (for 11 of 16 children). Target morphemes were then assigned with secondary priority to achieve balance of morphemes serving as targets and controls across children, and to allow for replication for the at least some of the same targets across children. Treatment and control morphemes were balanced across children and conditions to the extent possible, which reduced the likelihood that a difference in treatment effects across morphemes assigned to children was due to differences in the difficulty of acquisition of any one morpheme over the other. Refer to Table 2 for morpheme assignment information.

*Treatment Conditions.* Children were randomly assigned to one of two treatment schedule conditions: either the massed or spaced condition. Children in the massed condition ($n=8$) were provided treatment in a single 30-minute session. Children in the spaced condition ($n=8$) were provided treatment in three 10-minute sessions per day, each separated by 75 to 115 minutes, over a 3.5-hour period. The only difference between treatment conditions was whether the treatment occurred in three spaced (10-minute) sessions or one massed (30-minute) session each day of treatment during the intervention program. All other treatment parameters were held constant across the two conditions.
Table 2. Treatment Variables for Children in Spaced and Massed Conditions

<table>
<thead>
<tr>
<th>Participant</th>
<th>Target</th>
<th>Control</th>
<th>Tx Days</th>
<th>Participant</th>
<th>Target</th>
<th>Control</th>
<th>Tx Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>402</td>
<td>doesn’t</td>
<td>Wh-ques</td>
<td>26</td>
<td>203</td>
<td>she</td>
<td>doesn’t</td>
<td>23</td>
</tr>
<tr>
<td>201</td>
<td>she</td>
<td>Wh-ques</td>
<td>25</td>
<td>304</td>
<td>has</td>
<td>past</td>
<td>26</td>
</tr>
<tr>
<td>202</td>
<td>has</td>
<td>past</td>
<td>24</td>
<td>502</td>
<td>is-ing</td>
<td>past</td>
<td>26</td>
</tr>
<tr>
<td>301</td>
<td>is-ing</td>
<td>3ps</td>
<td>21</td>
<td>204</td>
<td>past</td>
<td>are x-ing</td>
<td>25</td>
</tr>
<tr>
<td>401</td>
<td>is-ing</td>
<td>3ps</td>
<td>25</td>
<td>403</td>
<td>past</td>
<td>3ps</td>
<td>25</td>
</tr>
<tr>
<td>302</td>
<td>past</td>
<td>has</td>
<td>25</td>
<td>404</td>
<td>3ps</td>
<td>she</td>
<td>26</td>
</tr>
<tr>
<td>601</td>
<td>3ps</td>
<td>is-ing</td>
<td>24</td>
<td>602</td>
<td>past</td>
<td>3ps</td>
<td>21</td>
</tr>
<tr>
<td>501</td>
<td>is-ing</td>
<td>she</td>
<td>26</td>
<td>305</td>
<td>is-ing</td>
<td>3ps</td>
<td>26</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td>24.5</td>
<td></td>
<td></td>
<td></td>
<td>24.75</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
<td>1.83</td>
</tr>
</tbody>
</table>

Treatment Context. Children participated in a six-week intervention program that was incorporated into a half-day language enrichment preschool program. They did not receive language therapy outside of the intervention program during that time. Treatment was provided in one-on-one sessions by one of five trained research clinicians. Each clinician provided treatment to the same children, either two children each (for two clinicians) or four children each (for three clinicians), throughout the intervention program. Clinicians provided treatment to an equal number of children in both experimental conditions (massed and spaced).

Treatment Procedures. Conversational recast is commonly used in interventions designed to help children acquire proper morpho-syntactic structure. In
In this study, focused recasts were used, which only target a single morpheme throughout the treatment period. In this procedure, the child and the clinician engage in child-directed play, and the clinician provides opportunities for the child to communicate. When the child makes a statement, called a “platform utterance” that should, could, or did contain an attempt at the target morpheme, the clinician restates the child’s utterance with the target morpheme used correctly. This provides an immediate, correct model of the target morpheme relevant to the child’s interest at that time. The clinician model (the recast) is always grammatically correct, but need not be corrective (i.e., recasts can follow correct child platform utterances). Recasts that follow either a correctly produced utterance or an incorrect utterance each facilitate acquisition of the target form (Hassink & Leonard, 2010). As a child makes progress on his or her treatment target, a clinician increasingly provides recasts that follow correctly produced utterances that include the treatment morpheme. All recasts were contingent upon a child’s preceding utterance and clinicians were not permitted to model the target grammatical morpheme outside of the actual recast.

A previous study by Plante et al. (2014) provided evidence for a version of conversational recast treatment, referred to as Enhanced Conversational Recast, in which the clinician incorporates high variability in the linguistic input provided to the child, supported by the incorporation of a wide variety of activities and attentional cueing. In the Plante et al. study, this was achieved by providing recasts that contained 24 different verbs in utterances containing target pronoun or verb morphology. When the target was a bound morpheme verb inflection (i.e., regular past tense –ed, 3rd person singular –s, auxiliary is [verb]—ing) or pronoun (i.e., she), 24 different verbs were used with the target in the recasts. When the target was the proper use of other
free morphemes (i.e., *has* and *doesn’t*), the different recasts contained unique lexical items surrounding the target, as appropriate (e.g., *It doesn’t fit; she doesn’t eat that; this one doesn’t have a tail*). Clinicians were instructed not to repeat other words in multiple recasts (e.g., *He rung there. He jumps there. He looks there*, when the target morpheme was a verb marker). Clinicians were not constrained in their use of recasts with regard to positions of the target morpheme within the utterance, utterance length, or vocabulary use (other than not using probe words and control morphemes, and not repeating the same lexical items frequently). Therefore, clinicians provided varied linguistic input in multiple ways.

Variability was further facilitated by directing clinicians to use a variety of materials and activities within and across session, increasing the unique contexts for morpheme use. Clinicians were required to use three activities per day as the context for conversational recasts. In addition, clinicians did not repeat activities frequently and always provided different activities on consecutive days. Some of the more common activities utilized include interactive book reading, craft activities, structured games, pretend play, and object play. These procedures resulted in the use of a large number of contexts and lexical items used with the target morphemes.

In order to facilitate lexical variability in the child’s speech and subsequent recasts, clinicians were encouraged to model verbs that could accompany the target morpheme. Usually, models were provided by using a verb’s uninflected form during play. This reduced the burden on the child to generate appropriate vocabulary that would provide the opportunity for the child to attempt the target morpheme. However, clinicians were discouraged from modeling verbs so much that it became pragmatically odd. To avoid excessive models of the root verb and allow sufficient wait time for
children to respond to prompts, a clinician would frequently provide the verb or relevant vocabulary and an opportunity for a child to use it in a play context. If the child did not attempt an utterance using the verb, the clinician could circle back to that lexical context later in play to give the child another opportunity to use target morphology with that or another verb. This was meant to provide a strong linguistic context for child utterances that include an attempt at the target grammatical morpheme.

Initially, some children provided little verbal output, limiting the clinician’s ability to provide sufficient recasts. For these children, in addition to modeling vocabulary the child could use, clinicians encouraged verbal output by eliciting or prompting platform utterances. Common prompts used by clinicians included direct requests, such as “Tell me what’s happening” or “What happened?” or “Tell me what the animal does when he is hungry”, and indirect prompts such as, “I wonder what is going on”. A study of features of conversational recasting that contributed to treatment outcomes indicated that providing recasts following elicited platform utterances and spontaneous platform utterances were equally effective (Hassink & Leonard, 2010). While recasts could occur following platform utterances that were not obligatory uses of the morpheme, obligated contexts gave children the best opportunity to use morphemes. Therefore clinicians were urged to use prompting techniques that made the use of the morpheme obligatory.

Clinicians were encouraged to maintain naturalistic interactions as much as possible, even if they needed to elicit most child utterances. For example, clinicians were told to follow the child’s lead and their focus during play in terms of the words they used. For children who provided minimal verbal output, clinicians were also
instructed to limit asking questions when not eliciting platform utterances, to reduce the likelihood that children would be overwhelmed by demands for information beyond what was needed for successful conversational recasting.

A critical feature of Enhanced Conversational Recast is that the child’s attention must be on the clinician at the time the recast is provided. The salience of the recasts is increased by providing attentional cues to the child, to assure attention at the time of the recast. These include saying the child’s name, lightly touching the child’s shoulder, or getting in the child’s visual field immediately before recasting the utterance. The choice of attentional cue was dictated by which type of cue a child reliably responded to. These cues were used if the child is not already watching the clinician prior to the recast. Clinicians received specific training for obtaining a child’s attention for the recast for children who did not consistently make eye contact during the conversational recast with minimal cueing. Therefore, while the treatment occurred in a naturalistic play context, there were attentional cues to look at the clinician. In an effort to increase a child’s attention during treatment in this study, the clinicians were instructed to follow the child’s lead with play activities and encourage participation by giving choices. Additionally, environmental planning allowed the clinicians to control the pace of the play interaction and to limit toys and play objects to enhance attending to the linguistic targets modeled by the clinician.

*Treatment parameters.* Treatment in this intervention program was provided using Enhanced Conversational Recast procedures in a one-on-one session (Plante et al., 2014). Accordingly, treatment *dose form* in this study consisted of a recast that used vocabulary that was unique to that recast and was administered to a child who was attending (i.e., looking at the clinician) during the recast, and these took place in a
one-on-one treatment session. The treatment dose was 24 conversational recasts per day targeting a specific grammatical morpheme, regardless of whether these were administered in the massed or spaced condition.

Other elements of dose include session duration, and the rate and distribution of doses over the treatment session (Warren, Fey, Yoder, 2007). The session duration was either 10 minutes (spaced) or 30 minutes (massed), and the dose frequency per day was either three sessions or one session, respectively. The duration x dose frequency yields daily intensity, which was equal across groups: 10 minutes x 3 sessions = 30 minutes or 30 minutes x 1 session = 30 minutes. Thus, children in both groups received a total of 24 conversational recasts in 30 minutes of therapy per day. The weekly dose frequency for all children was 5 times per week with the exception of occasional child absences.

The overall rate of delivery was controlled across the spaced and massed conditions at 8 recasts per each 10-minute block of time. Clinicians were also instructed to distribute recasts as evenly as reasonably as possible throughout the sessions, though the actual distribution of recasts within each 10-minute interval was not monitored. Clinicians were given timers set to 10-minute intervals to facilitate their adherence to the 8 recasts per 10 minutes rate of delivery. This rate of delivery averaged to approximately one recast every 1.25 minutes.

The total intervention duration was just over 5 weeks. The number of days of the intervention possible was 26, and children attended approximately 25 days on average (M = 24.625, range=21-26 days). Five children (202, 301, 305, 203, 602) were absent three or more days during the intervention program. Three of these children (in bold) attended two additional individual language therapy sessions on
two days during the week following the end of the program. A t-test confirmed that the number of treatment days children attended did not differ in the spaced condition \( (M = 24.5, SD = 1.6) \) and massed condition \( (M = 24.75, SD = 1.8) \), \( t = 0.290, p = 0.776, d = .136 \). See Table 2 for treatment days attended by children in spaced and massed conditions.

The average of twenty-five days of intervention yields a cumulative intervention intensity of 600 conversational recasts containing the target morpheme (24 recasts per day for 25 days) provided in 750 minutes of treatment (30 minutes per day for 25 days). The actual ranges for participants involved in this study were 21-26 days; 504-624 recasts; and 630-780 minutes.

Generalization Probes. In order to assess each child’s progress during the five-week treatment period, generalization probes were administered three days per week. The three times per week (rather than daily) probe schedule was intended to reduce the likelihood that children would get into a “set”, or a non-naturalistic response pattern that may result from very frequent probing. Generalization probes were always conducted prior to the administration of any treatment, regardless of whether treatment was massed or spaced. This was intended to assure that generalization probe data would reflect a child’s previous learning, and not learning from earlier that same day.

Generalization probes were conducted using a set of verbs and materials included in thematic toy sets called probe kits (e.g., sea world, soccer players) that were never used during treatment. There were six different probe kits so that clinicians could utilize a variety of probe kits throughout the five weeks. Clinicians were instructed not to repeat probe kits during a week, and to use every kit at least
once during the five weeks of treatment. Clinicians were free to choose whether to use a single probe kit or two probe kits to elicit both target and control morphemes.

A set of 20 verbs that were not used during treatment were reserved for use during generalization probe sessions. Clinicians were instructed to use a subset of 10 verbs, of the total set of 20 verbs for probing both the target and control morphemes. However, it was not necessary to use the same 10 verbs to probe both the target and control morphemes. The use of verbs (or lexical items) reserved for generalization probes assured that a child's performance on these probes reflected generalization, rather than prior experience with the toy sets or particular vocabulary.

To begin a probe activity, the clinician established a realistic play or story telling context that was appropriate to a particular probe kit (e.g., watching a toy character feed or take care of animals, or talking about what happened in a soccer game). When clinicians probed for verb tense-markers, the clinician set a context to establish the appropriate tense of the actions without modeling the target form. For example, the clinician may have said, “Let’s watch Ana [a toy girl] feed the animals. It’s her job. What does Ana do?” to establish a context to elicit the use of third-person singular s, or use of the pronoun she. The clinician demonstrated an action and either continued the action (e.g., for present progressive tense) or stopped the action (e.g., for past tense) while also modeling probe vocabulary item stem, and then probed for use of the morpheme (e.g., for past tense: completing an action and then asking, “What happened with the tiger?”). The clinician provided ten obligatory contexts for the use of each target morpheme as well as ten obligatory contexts to probe for the control morpheme in each probe session.
Long-Term Retention. A follow-up appointment was planned to take place approximately 8 to 10 weeks after the end of the treatment program in order to obtain a measure of long-term retention. During the follow-up appointment, generalization probes were administered to probe for the child’s use of the treatment and control morphemes. These probes were identical to those used during the treatment program in terms of the use of probe kits and probe words, obligatory contexts, and individual administration. The person conducting the probe and the treatment room were both familiar from the preschool language program. However, the person conducting the follow-up probe session was not the clinician who treated the child during the treatment phase.

Experimental Control

Parents and classroom teachers were blind to which element of grammar served as the Target and Control morphemes for each child. Although informed that their child would be receiving treatment, parents were not present to observe treatment sessions. They were provided with weekly newsletters concerning the classroom curriculum of the preschool program and were given a summary of their child’s progress after the conclusion of the program. Clinicians were not told about the study hypothesis or theoretical framework driving the hypothesis.

Fidelity & Reliability

Procedural adherence (fidelity) and reliability in recording children's oral responses are necessary components of well-controlled treatment studies (Kaderavek & Justice, 2010). Both reliability and fidelity were established for generalization probes and treatment sessions through several steps. In addition, all sessions were
audio-and video-recorded and clinicians were free to consult the videos to resolve any questions they may have had about child utterances.

Clinician Training. The five research clinicians who provided individual language therapy to children were undergraduate or graduate students in speech-language pathology. They provided treatment under the supervision of two certified speech-language pathologists. Clinicians were trained for two half days (eight hours total) to facilitate a high level of treatment fidelity. Training for clinicians included reviewing the fundamental components of conversational recast through reading, discussion, video review, practice, and feedback.

Fidelity Tracking. To facilitate fidelity of treatment administration, clinicians were provided with clipboards and data collection sheets that they could use to track the number and lexical diversity of the recasts provided as they were delivered during treatment. The clinician also recorded whether the child’s attempt (the platform utterance) had been correct or incorrect, and any spontaneous use of the target morpheme occurred. Clinicians also received daily feedback during the first two weeks of the intervention program to ensure that treatment requirements were being met. Feedback concerning adherence to the treatment protocol was provided in an ongoing basis as needed throughout the treatment period. Treatment fidelity was evaluated live for 11% of treatment sessions (n=47). Because not all of the fidelity components were tracked in every session, the base rate for each component ranged between 32 and 47 sessions. For a session to be conducted with 100% fidelity in all respects, the following criterion must have been satisfied:

1) the clinician used

   (a) at least three different activities per day,
(b) non-probe kit activities

2) the clinician provided
   (a) 24 properly administered recasts,
   (b) recasts that included 24 unique verbs
   (c) recasts that were provided at a rate of 8 recasts every 10 minutes

3) Clinicians avoided the use of probe words.

Deviations from aspects of fidelity were distributed across the five clinicians, with no one clinician representing a disproportionate number of deviations.

In six of 44 daily sessions, only two activities were used in therapy during a given day. The resulting treatment fidelity corresponding to 1(a) was 86.36%. No probe kit activities were used in any treatment session. Treatment fidelity corresponding to 1(b) was 100%.

All 47 treatment sessions evaluated for fidelity involved at least 24 properly administered recasts incorporating 24 different verbs. By this definition, the treatment fidelity score corresponding to 2(a) and 2(b) was 100%. In 15 sessions (nearly 30%), clinicians provided more than 24 total recasts. In 13 cases, the clinician provided one extra recast; in one case, the clinician provided 2 extra recasts, and in one case, the clinician inadvertently provided 5 extra recasts. Because 1148 recasts were provided over the 47 sessions instead of 1128 recasts, this resulted in a total difference of 20 extra recasts, or 1.77% additional recasts. In the vast majority of these instances, the clinician provided an extra recast because they had inadvertently repeated a lexical item (e.g., verb) in the 24 recasts, and needed to provide one more recast to include 24 unique lexical items in recasts. This deviation from strict adherence to 24 recasts was mostly due to the clinician responding appropriately (i.e.,
recasting) to child-initiated utterances, and was not considered to have any negative effect on treatment implementation.

*Spontaneous use* of the target morpheme also occurred during treatment, defined by when the clinician did not provide a verb or vocabulary to solicit a platform utterance from the child. Most of the instances in which a child spontaneously used a target morpheme correctly, the clinician provided a recast of the child’s utterance. Occasionally, an utterance could not be recasted because it contained a probe word or it had been previously recasted that day during treatment (including cases when the child used the same word with the target morpheme more than one time).

Rate was evaluated during 32 treatment sessions. There were two violations to the rate of 8 recasts every 10 minutes, and each time was due to a child factor such as needing a bathroom break or needing a break for behavioral support. Therefore, treatment fidelity for the *rate* parameter, 2(c), was 93.75%.

A probe word was used only one time in one session of the 47 evaluated for treatment fidelity. In that case, after realizing that a probe word had been used, the clinician provided one additional recast (total of 25 recasts) to ensure 24 unique, non-probe word recasts were provided. Treatment fidelity corresponding to number 3 above was 98%; 46 of 47 treatment sessions did not contain the use of a probe word.
Generalization Probe Procedural Fidelity. For probe sessions, fidelity was determined based on two primary components: setting up ten (10) obligatory contexts each to probe for use of target and control morphemes, and using probe kit materials and words. If the clinician did not provide an obligatory context, a subsequent incorrect attempt by the child was not counted as a probe response. On the rare occasion this occurred, the child was given another opportunity to use the probed morpheme in an obligatory context. Conversely, subsequent use of the target following a non-obligatory context was counted as a correct use during a probe session.

Clinicians recorded data on the three components of fidelity (obligatory contexts, and probe kit used and probe words uses) while administering the probes and recording the child’s responses. No clinician ever used materials other than those from the probe kits during probe sessions, and probe kit materials were never used during treatment sessions by any clinician.

Reliability of Probe Data. Reliability in recording the oral responses of children during generalization probes was confirmed by a second data tracker in the room who

---

**TABLE 3. Treatment Fidelity: Scores by Treatment Fidelity Components**

<table>
<thead>
<tr>
<th>Fidelity Component</th>
<th>Fidelity Score</th>
<th>Evaluated in n sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(a) 3 different activities</td>
<td>86%</td>
<td>44</td>
</tr>
<tr>
<td>1(b) not probe kit activities</td>
<td>100%</td>
<td>47</td>
</tr>
<tr>
<td>2(a) 24 proper recasts</td>
<td>100%</td>
<td>47</td>
</tr>
<tr>
<td>2(b) 24 unique lexical items</td>
<td>100%</td>
<td>47</td>
</tr>
<tr>
<td>2(c) Rate of recasts</td>
<td>94%</td>
<td>32</td>
</tr>
<tr>
<td>3 No probe words used</td>
<td>98%</td>
<td>47</td>
</tr>
</tbody>
</table>
recorded a child’s responses to all probed structures in 12.1% (n=31) of generalization probe sessions. Data trackers were trained and demonstrated the ability to carefully listen to and record a child’s responses while simultaneously evaluating fidelity of probe administration. Data trackers responsible for tracking probe reliability were blind to which morpheme was the child’s target, and which was the control. Reliability was based on point-to-point agreement for ten probes for the target and ten probes for the control, for a total of 20 data points in each generalization probe session. Additionally, both the clinician and the data tracker recorded spontaneous uses of the target morpheme, though this rarely occurred and is therefore not included in the reliability calculation. Reliability calculated for 31 generalization probe sessions ranged from 85%-100%, with an average of 96%. Reliability coding occurred for two generalization probe sessions for each child, including one of the three end-treatment generalization probes used to calculate each child’s $d$ scores (except for 602 for whom only one generalization probe session was coded for reliability due to absence during the final week of the program). The reliability averages calculated by clinician fell in the range of 93%-100%, indicating strong agreement across clinicians and data trackers. These averages indicate a high level of reliability in recording children’s oral responses (attempts of productions of target and control morphemes) during generalization probe sessions.

Results

For descriptive purposes, effect size $d$ was calculated to reflect change for each child’s target and control morphemes. These are reported in Table 4. The graphical
representation of effects for Target and Control morphemes for both massed and spaced conditions is presented in Figure 1.

Effect size is an expression of change in performance relative to starting performance. There are various ways to calculate effect size, but for the purpose of this study, the following method, originally used in Plante et al. (2014) was utilized. The treatment effect size $d$ was calculated by subtracting mean correct morpheme use during the first three pretreatment probes (baseline) from mean correct morpheme use on final three generalization probes (obtained in the last week of treatment). This difference was divided by the standard deviation of the final 3 generalization probes. In cases when there was no variance in the final three generalization probes (they were all the same value), then the minimum possible standard deviation value was used (a difference of 1 response (10%) among the three days). This is equivalent to an approximate value $= 0.577$ for the standard deviation. This method is conservative statistically in that it inflates the denominator slightly (reducing the effect size slightly), but prevents dividing the numerator by zero in cases where there is no variance. Additionally, the use of the standard deviation of the final three generalization probes, rather than the pretreatment or pooled pretreatment and end treatment standard deviation, is also a conservative estimate of $d$. Because children with SLI often show no pretreatment use of target morphemes, zero variance in the pretreatment probes would lead to a very small denominator and an inflated $d$ value.
Figure 1. Average effect size for Target and Control morphemes by Massed and Spaced treatment condition. Error bars represent the standard errors.
Table 4. Performance data for children in the Spaced and Massed treatment conditions.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Target Form</th>
<th>Control Form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target</td>
<td>Pre Tx</td>
</tr>
<tr>
<td>402</td>
<td>doesn’t</td>
<td>0%</td>
</tr>
<tr>
<td>201</td>
<td>she</td>
<td>3%</td>
</tr>
<tr>
<td>202</td>
<td>has</td>
<td>3%</td>
</tr>
<tr>
<td>301</td>
<td>is -ing</td>
<td>0%</td>
</tr>
<tr>
<td>401</td>
<td>is -ing</td>
<td>0%</td>
</tr>
<tr>
<td>302</td>
<td>past</td>
<td>10%</td>
</tr>
<tr>
<td>601</td>
<td>3ps</td>
<td>7%</td>
</tr>
<tr>
<td>501</td>
<td>is -ing</td>
<td>20%</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>5.38%</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>6.93%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant</th>
<th>Target Form</th>
<th>Control Form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target</td>
<td>Pre Tx</td>
</tr>
<tr>
<td>203</td>
<td>she</td>
<td>0%</td>
</tr>
<tr>
<td>304</td>
<td>has</td>
<td>3%</td>
</tr>
<tr>
<td>502</td>
<td>is -ing</td>
<td>3%</td>
</tr>
<tr>
<td>204</td>
<td>past –ed</td>
<td>16%</td>
</tr>
<tr>
<td>403</td>
<td>past –ed</td>
<td>16%</td>
</tr>
<tr>
<td>404</td>
<td>3ps</td>
<td>33%</td>
</tr>
<tr>
<td>602</td>
<td>past –ed</td>
<td>3%</td>
</tr>
<tr>
<td>305</td>
<td>is –ing</td>
<td>3%</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>9.63%</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>11.31%</td>
</tr>
</tbody>
</table>
Generalization Probes

The primary outcome measure was the child’s use of target and control morphemes during a play-based probe that obligated the use of the morphemes with untreated lexical items (the generalization probes). Analyses of the primary outcome measure were based on the specific study hypotheses.

Effect of treatment. Based on previous demonstrations of effectiveness of Enhanced Conversational Recast treatment, one overarching hypothesis was that this treatment would result in statistically significant change in target morpheme production during the treatment period, without equivalent gains on the control morpheme. This should be true regardless of the condition of treatment delivery schedule (Spaced or Massed). This hypothesis was tested by comparing performance (mean percent correct) on the final three probes occurring the last week of treatment to performance on the three pretreatment probes. Consistent with this hypothesis, children used their target morphemes significantly more frequently during the final three probe sessions than during the three pretreatment probe sessions $t(15) = 5.65$, two tailed $p < .0001$, $d = 1.241$. This significant increase applied to both the Massed delivery condition $t(7) = 4.849$, $p = .0019$, $d = 1.344$, and the Spaced delivery condition $t(7) = 3.25$, $p = .014$, $d = 1.123$. The increase in morpheme use could not be accounted for by maturation or the general language stimulation that children received in the accompanying preschool language program, given that the control morphemes did not show a similar gain from pretreatment probes to end-treatment probes $t(15) = .031$, $p = .976$; $d = -.0049$. Therefore, the change in morpheme use for the treated morpheme can be considered a treatment effect. Overall treatment efficacy was demonstrated by an effect size $d$ for treated morphemes significantly greater than the effect size $d$ for
untreated morphemes for participants in both conditions $t(14) = 4.5549, p = 0.0002, d = .952$.

*Effect of Massed vs. Spaced Dose Schedule.* The purpose of the experimental manipulation in this study was to determine if either massed or spaced dose schedule of Enhanced Conversational Recast therapy would lead to a greater treatment effect. Thus, analysis of the primary outcome measure was limited to tests of the specific study hypotheses relating to the experimental manipulation of dose schedule. This targeted approach avoided correction for comparisons that were not of interest, but would still be included in an omnibus analysis (e.g., ANOVA).

This was tested using a hypotheses-driven comparison (two-tailed t-test). An independent samples t-test was used to test for differences in end treatment use of the target morphemes for each group. The end treatment percent correct use did not differ between conditions, $t(14)= .436, p = .669, d = .057$, failing to confirm the hypothesis of a difference between treatment delivery methods. In addition, the effect size for this comparison was very small, suggesting that a difference would be unlikely even with a much larger sample size. A power analysis indicated that 4,900 subjects would be required to achieve 80% power with this effect size.

*Control Analyses.* In addition to the hypothesis-driven analyses described above, control analyses were used to rule out alternative explanations. The following t-tests were used to rule out other explanations for the main results:

1. Differences in pretreatment use of the treated morphemes.
2. Correlations between language scores (SPELT-P2, PPVT, TOLD-GU) and nonverbal IQ scores (K-ABC) and treatment effect size $d$ for treated morphemes.
Children did have a range of pretreatment use of their target morphemes, but the spaced and massed groups did not differ significantly in terms of pretreatment use $t(14) = .906, p = .380, d = .376$. Recall also that the d-statistic data reported in Table 4 and Figure 1 incorporates the pretreatment use data by reflecting the magnitude of change achieved during the treatment period.

Correlation data were calculated to determine if standardized test scores related to treatment outcomes (target effect size d). All reported correlations were tested for significance at $p < .05$. Pearson’s $r$ was non-significant for SPELT-2, $r(14) = .276$; TOLD-GU, $r(14) = .356$; and GFTA-2, $r(14) = .181$. The PPVT-4, $r(14) = .490$ was significantly correlated with treatment effect size d. There was also a significant correlation of K-ABC scores to treatment effect sizes, $r(14) = .599$. The K-ABC was also evaluated for co-variance with the language measures. The correlation of K-ABC and SPELT-P2 was not significant (n.s., $r(14) = .340$). There were significant correlations of K-ABC and PPVT-4, $r(14) = .524$; and K-ABC and TOLD-GU, $r(14) = .509$.

In-Treatment Performance

For most children, the increase in target morpheme use during probe sessions (a measure of generalization) paralleled use of the target morphemes during treatment sessions. As in generalization probes, the clinician typically provided the verb stem or other platform utterance vocabulary when eliciting child utterances to recast during treatment sessions, unless children spontaneously produced a statement using or requiring the use of the target morpheme. The in-treatment use of target morphemes is presented in Table 5. An independent samples t-test was used to test for group
differences in the in-treatment productions of the target morpheme. As with
generalization probe responses, in-treatment target use did not differ significantly
between spaced and massed conditions \( t(14) = .179, p = .861, d = .084 \). In-treatment
target morpheme use was significantly correlated with treatment effect size \( d = .616, p < .05 \).

Spontaneous Use

Recall that use of the target morpheme by the child was considered
spontaneous when the child used a verb or lexical item not previously provided
verbally by the clinician during that session. Spontaneous use of the target morpheme
during treatment is reported in Table 5. While there was a fairly wide range of
spontaneous use across children, spontaneous use did not differ significantly between
spaced or massed treatment conditions \( t(14) = .862, p = .403, d = .383 \).

Figure 2. Spontaneous use of target morpheme during treatment sessions for spaced
and massed conditions. Error bars represent the standard errors.
### Table 5. In-Treatment target morpheme use & Spontaneous use for children in Spaced and Massed Conditions

<table>
<thead>
<tr>
<th>Participant</th>
<th>In-Tx Use</th>
<th>Spontaneous use</th>
<th>Participant</th>
<th>In-Tx Use</th>
<th>Spontaneous use</th>
</tr>
</thead>
<tbody>
<tr>
<td>402</td>
<td>457</td>
<td>7</td>
<td>203</td>
<td>196</td>
<td>3</td>
</tr>
<tr>
<td>201</td>
<td>165</td>
<td>8</td>
<td>304</td>
<td>278</td>
<td>11</td>
</tr>
<tr>
<td>202</td>
<td>237</td>
<td>3</td>
<td>502</td>
<td>234</td>
<td>37</td>
</tr>
<tr>
<td>301</td>
<td>173</td>
<td>13</td>
<td>204</td>
<td>384</td>
<td>19</td>
</tr>
<tr>
<td>401</td>
<td>236</td>
<td>6</td>
<td>403</td>
<td>135</td>
<td>6</td>
</tr>
<tr>
<td>302</td>
<td>141</td>
<td>31</td>
<td>404</td>
<td>260</td>
<td>26</td>
</tr>
<tr>
<td>601</td>
<td>25</td>
<td>6</td>
<td>602</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>501</td>
<td>35</td>
<td>1</td>
<td>305</td>
<td>44</td>
<td>1</td>
</tr>
</tbody>
</table>

| Mean        | 183.63    | 9.38            | Mean        | 195.13    | 14.13           |
| SD          | 136.20    | 9.43            | SD          | 120.70    | 12.4            |

### Individual Results

Data from individual participants collected from the three-times weekly generalization probes for each child demonstrate how the overall results of the treatment are reflected in the performance of individual children. These plots also provide additional information, including the trajectory of learning over time, and time-to-response to treatment. The data were plotted to allow for both visual inspection of trends, and the ability to obtain additional qualitative comparisons, particularly for the children who reached ceiling performance.

The graphs provide information on the time frame needed for children to reach...
ceiling performance. One child (402) reached ceiling performance (100%) on the 5th probe day, but did not solidify ceiling performance until the tenth probe session. One child (304) reached and maintained ceiling performance (i.e., respond correctly to every probe for the treatment morpheme; scored 100%) after 15 treatment sessions (on the ninth probe session during the treatment period). Two other children reached ceiling performance and maintained ceiling or near ceiling performance (90%-100%) after 17 (402) and 19 (502) treatment days, (on the tenth and eleventh probe days, respectively). After 21 treatment days, three more children (201, 202, 203) achieved and maintained ceiling or near ceiling performance. Although increases in performance as the program progressed are evident for some of the remaining ten children, eight of these children did not reach ceiling or near ceiling performance during the treatment period. One child (204) scored 100% in one probe session, but not again in any of the several subsequent probe sessions. Another child (403) scored 100% in the final probe session only, after he had received 24 days of treatment.

Figure 3. Individual subject data.
Children are grouped based on their treatment response. Children who showed ceiling performance (10 correct probe responses out of 10 possible responses) over multiple probe sessions before the end of the treatment period AND achieved an effect size of 15.01 or greater were classified as “Excellent Responders”. Those with strong, but not ceiling performance and achieved an effect size of 2.18-8.0 were classified as “Good Responders”. Those who showed weak performance and achieved an effect size of 1.01 or smaller are classified as “Poor Responders”. The grey dashed line separates the pretreatment and treatment periods. T: treated morpheme C: Control morpheme. PRE 1-3: Pretreatment probes 1 through 3. GEN 1-12: Generalization probes 1 through 12. END 1-3: End treatment probes 1-3.
Excellent Responders (n=6)

Spaced

402; d=17.32

Massed

203; d=16.74
### Good Responders (n=5)

#### Spaced

<table>
<thead>
<tr>
<th>Interval</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>301</td>
<td>8.0</td>
</tr>
<tr>
<td>401</td>
<td>5.02</td>
</tr>
<tr>
<td>404</td>
<td>2.18</td>
</tr>
</tbody>
</table>

#### Massed

<table>
<thead>
<tr>
<th>Interval</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>204</td>
<td>5.33</td>
</tr>
<tr>
<td>403</td>
<td>4.37</td>
</tr>
</tbody>
</table>

**Graphs:**
- **Spaced:**
  - T: aux is -ing
  - C: 3rd pers -s

- **Massed:**
  - T: past -ed
  - C: aux are -ing
Poor Responders (n=5)

Spaced

302, d=1.01

- T: past -ed
- C: has

Massed

602, d=0.66

- T: past -ed
- C: 3rd pers -s

601; d= -0.58

- T: 3rd pers -s
- C: aux. is -ing

305; d= 0.62

- T: aux is -ing
- C: 3rd pers -s

501; d=-3.46

- T: aux is -ing
- C: she
Long-term retention

In total, thirteen of the sixteen children enrolled in the summer treatment program returned for a follow-up measure in the early fall, between seven and eleven weeks following the end of the treatment program. Family availability was the only factor that influenced the scheduling of the follow-up appointment. Three children (203, 501, 601) did not return for follow-up testing because they were not available due to family schedule or the investigator was not able to make contact with the families.

The percent correct use of Treated and Control morphemes at follow-up is reported in Table 6. A two-tailed t-test indicated that there was no difference between performance on the target morpheme on the follow-up (retention) measure for massed and spaced groups, \( t(11) = .0783, p = .939, d = .042 \). For descriptive purposes, positive retention was defined as scoring no more than 20% lower on the follow-up measure than the average score on the final three generalization probes, and having a score of at least 50% correct on the follow-up measure. According to this criteria, nine of the thirteen children (4 in the spaced condition and 5 in the massed condition) who returned for follow-up testing demonstrated positive retention of learning with respect to end-treatment performance.
### Table 6. Performance: End-Treatment and Long-term Retention

#### Spaced Condition

<table>
<thead>
<tr>
<th>Participant</th>
<th>Target Form</th>
<th>Control Form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>End Tx</td>
<td>Follow Up</td>
</tr>
<tr>
<td>402</td>
<td>doesn’t 100%</td>
<td>90%</td>
</tr>
<tr>
<td>201</td>
<td>she 100%</td>
<td>100%</td>
</tr>
<tr>
<td>202</td>
<td>has 93%</td>
<td>90%</td>
</tr>
<tr>
<td>301</td>
<td>is -ing 80%</td>
<td>40%</td>
</tr>
<tr>
<td>401</td>
<td>is -ing 77%</td>
<td>80%</td>
</tr>
<tr>
<td>302</td>
<td>past 33%</td>
<td>20%</td>
</tr>
<tr>
<td>601</td>
<td>3ps 3%</td>
<td>0%</td>
</tr>
<tr>
<td>501</td>
<td>is -ing 0%</td>
<td>DNR</td>
</tr>
</tbody>
</table>

| Mean        | 60.63%      | 60%          | 7.80% | 18.57% |
| SD          | 42%         | 39.48%       | 9.99% | 20.35% |

#### Massed Condition

<table>
<thead>
<tr>
<th>Participant</th>
<th>Target Form</th>
<th>Control Form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>End Tx</td>
<td>Follow Up</td>
</tr>
<tr>
<td>203</td>
<td>she 96%</td>
<td>DNR</td>
</tr>
<tr>
<td>304</td>
<td>has 100%</td>
<td>100%</td>
</tr>
<tr>
<td>502</td>
<td>is -ing 97%</td>
<td>90%</td>
</tr>
<tr>
<td>204</td>
<td>past –ed 70%</td>
<td>50%</td>
</tr>
<tr>
<td>403</td>
<td>past –ed 83%</td>
<td>0%</td>
</tr>
<tr>
<td>404</td>
<td>3ps 66%</td>
<td>80%</td>
</tr>
<tr>
<td>602</td>
<td>past –ed 13%</td>
<td>DNR</td>
</tr>
<tr>
<td>305</td>
<td>is –ing 27%</td>
<td>50%</td>
</tr>
</tbody>
</table>

| Mean        | 68.88%      | 61.67%       | 8.63%  | 25%         |
| SD          | 32.79%      | 36.56%       | 18.25% | 21.68%      |

DNR= Did not return for follow-up testing.

Bolded scores in the follow-up column indicate the criteria for positive retention were met.
Discussion

The first hypothesis, that Enhanced Conversational Recast would be efficacious in that it would produce statistically significant change in target morpheme production during the treatment period, without equivalent gains on the control morpheme, was confirmed. Additionally, follow-up probes collected approximately two months later provide evidence of long-term retention of learning that was demonstrated during the treatment period. As such, this study serves as a replication of the efficacy of Enhanced Conversational Recast for grammatical morpheme acquisition for preschool age children with specific language impairment (SLI). The present outcomes are consistent with previous results for children who received high variability input during Enhanced Conversational Recast treatment (Plante et al., 2014). The effect size obtained in this study, $d = 1.241$, is similar to the effect size obtained in Plante et al., $d = 0.92$. Moreover, the treatment reported in this study was administered by clinicians who were not involved in any previous conversational recast study, and treatment was provided to a different cohort of preschool age children with SLI than had been studied previously. These factors further enhance the external validity of the Enhanced Conversational Recast approach.

The second hypothesis was that outcomes for Enhanced Conversational Recast treatment would differ based on whether the dose schedule involved massed or spaced dose delivery. The results indicated that there was no difference between groups. Furthermore, the very small effect size ($d = .057$) suggests that no difference would be present even with a much larger sample of children.
Given the wealth of experimental evidence to support the benefit of distributed learning (e.g., Janiszeweski et al., 2003), it was important to explore the potential of a spaced dose schedule to improve Enhanced Conversational Recast. However, treatment studies that have contrasted massed vs. spaced dose schedule have not clearly favored either massed or spaced dose delivery. The largely comparable outcome from spaced and massed dose delivery in the present study may be attributed primarily to two main causes. First, and most importantly, Enhanced Conversational Recast is a highly effective treatment. Therefore, the effect of the treatment technique likely overwhelmed any effect associated with differences in dose delivery. Had learning been less robust overall, it is possible that a dose schedule effect may have emerged. Consistent with this idea, Riches et al. (2005) showed the greatest advantage for spaced doses relative to massed doses for the more difficult-to-learn aspect of their task. In that study, children with SLI were asked to demonstrate both comprehension and production of the newly learned verbs. Although typically developing children in the study showed no difference in learning based on dose differences, children with SLI benefitted from higher frequency input (more exposures) in a more widely spaced dose (4 days versus 1 day) for the more difficult aspect of the verb learning task, verb production, for which overall performance was relatively poor.

Secondly, the current study involved a manipulation of dose schedule within a given day. Most other treatment studies investigating an intensity component manipulated rather the relative spacing of treatment over the course of days or weeks (e.g., Barratt et al., 1993; Smith-Lock et al., 2013a). The particular dose
schedules used in the current study assured that all children received the same number of doses and the same number of days of treatment over an equal time period (5 weeks). However, this treatment schedule raises the possibility that the role of daily treatment is contributing to the efficacy of this treatment, and that spacing of input within each day is not a critical factor. Other studies have found superior effects for spaced treatment provided fewer doses over more days in the spaced than massed conditions (e.g., Riches et al., 2005, Ukranitz et al. 2009).

Relationship to other studies. The current study is the only known experimental manipulation of dose schedule, while holding cumulative intervention intensity constant, for the treatment of morphosyntax in children with SLI. Therefore, direct comparisons to identical studies are not possible. Rather, the purpose of this discussion is to provide a general explanation of why dose spacing effects that occurred in previous studies and in other language domains are not evident in the current study. As this discussion will show, the complexity of dose schedule and frequency components, and their combination means that the effects of dosage manipulations alone are not straightforward.

Smith-Lock et al. (2013a) conducted one of the few studies of morphosyntax that considered dose schedule. Children with SLI received treatment for one hour per week for eight weeks (spaced) or treatment for one hour per day for eight days (massed). They reported greater improvement on their grammar elicitation task (Smith-Lock et al., 2013b) in the spaced condition. The cumulative intervention intensity in terms of treatment minutes was consistent for spaced and massed
groups. However, the treatment durations varied substantially (8 days versus 8
weeks) and the periods of time that elapsed from the end of treatment to
performance measurement were highly discrepant (6 weeks after treatment end in
the massed condition versus immediately after treatment end in the spaced
condition). Also, the authors reported that dose numbers were not measured and
may not have been equivalent across groups. Therefore, some caution is warranted
in terms of whether this study reflects simply the effect of massed vs. spaced
treatment.

There were a number of additional difficulties with how the Smith-Lock et al
(2013b) study measured treatment outcome. This study used a grammar elicitation
task to calculate “gain scores”. The manner in which the gain scores were calculated
and interpreted introduced a number of potential problems. Gain scores were
calculated based on a score out of 30 possible points on the grammar elicitation test.
The first problem is that gain scores were calculated for change in the pretreatment
baseline period, which was measured over 8 days or 8 weeks, depending on massed
or spaced group assignment, respectively. Measuring change over differing periods
of time may not be appropriate as performance has more opportunity to vary when
baseline is measured over relatively long periods of time.

Second, the “baseline” measures that were collected involved two data points
only, instead of a recommended minimum of three (Barlow & Hersen, 1984).
Although there were non-significant differences in pretreatment scores across
groups, there was a trend of higher scores for the 8 days (massed) group, and an
increasing trend between the two “baseline” data points obtained for that group.
This increasing trend suggests that a stable baseline was not obtained for the massed group. In contrast, there was no increasing trend in the two baseline data points for the 8 weeks group. The inconsistent time frame over which baseline was measured, and subsequent interpretation of gain scores obtained at these points is problematic. This suggests that the results of Smith-Lock et al., (2013a) may be more similar to the “no difference” result obtained in the present study than it may appear from their overall conclusion.

There have been a small number of studies that have examined word learning, rather than morpheme learning, using massed and spaced dose delivery schedules. Childers and Tomasello (2002) demonstrated a production advantage for newly learned words by very young typically-developing children (mean age= 2.5 years) who were given distributed exposures. This finding did not extend to grammar learning in children with SLI in the present study, who are approximately two to three years older. However, children with SLI were included in a later study by Riches et al. (2005) studied massed vs. spaced presentations of novel verbs to children with SLI. These children had a mean age very close to that of children in present study (5;6 vs. 5;4). The outcomes of this study were mixed. The younger, typically developing children showed no differential effect for massed vs. spaced presentations. The children with SLI showed an effect of distributed learning only on the production aspect of the task—the more difficult component of word learning compared to recognition or comprehension. This suggests that there may have been an effect of spacing in the present study, if the children had not performed so well overall. In addition, word learning differs from grammatical
learning in several potentially ways, which have been discussed previously. Either or both factors could explain why at no dose schedule effect was found for the present study, even when it has sometimes been found for word learning studies.

There is a large body of experimental evidence to show that widely-spaced exemplars facilitate verbal learning more so than massed exemplars in young children (e.g., Ambridge, Theakston, Lieven, & Tomasello, 2005; Riches, Tomasello, & Conti-Ramsden, 2005; Vlach et al., 2008). One commonality among these studies is that exposures to the target forms were limited (10-18 exposures) and provided over relatively few sessions (1-10 sessions). These dose and treatment intensities are in stark contrast to the current study in which approximately 600 exemplars were presented over 25 days. This difference suggests that the enhanced effect of distributed exposures may be more pronounced when the total number of exposures and sessions are low. This interpretation consistent with Childers and Tomasello (2002) who proposed that the most important dose-related factor for learning was the number of different days of exposure, with more days facilitating target word production. In terms of both schedule and intensity, the length of the training period appears to be a mediating factor.

These studies collectively lead to support for attending to several elements of dosage simultaneously. These have generally been over-looked, even in studies designed to investigate dosage, intensity, and schedule. It is likely that important considerations in the study of massed versus spaced treatment include the dose per unit of time, and the rate and distribution of these doses and the days over which the doses are administered. In addition, time between sessions is a potentially
important factor with relevance to the study of treatment schedule. The current study addressed the effect of incorporating sessions that were frequent, but differed in terms of dose distribution. However, these other factors have yet to be thoroughly explored alone or in combination.

Previous conversational recast studies that did not manipulate dose schedule have reported information that can be used to understand distribution effects on learning utilized in those studies. It is possible to draw basic conclusions about whether or not it is likely for learning to be facilitated given a particular treatment intensity and schedule. To make reasonable comparisons, studies of conversational recast targeting morphosyntax for children with SLI with similar, positive and significant effect sizes are considered. Leonard and colleagues (2006) reported treatment effect sizes of $d = .90$ for third person singular –s, and $d = 1.25$ for auxiliary is –ing. These average effect sizes are statistically significant and indicate that at least some children benefitted from the treatment. The average effect size for the high variability group (including a variety of bound morpheme targets) in Plante et al. was in the same range, at $d = .92$. The average effect size in the current study was $d = 1.24$, which is also in the range of effect sizes obtained in these studies.

The Leonard et al. studies (2004, 2006) involved four 30-minute treatment sessions per week, condensed into two lab visits per week for six months. This means that children received a treatment session, followed by a general language stimulation activity, followed by another treatment session on 48 different days. This schedule represents both massed and spaced components, though the study was not designed to address schedule or intensity. Nevertheless, at least some
children were able to demonstrate a treatment effect given this treatment schedule and duration. However, effect sizes obtained after six months of treatment (Leonard et al., 2006) were equivalent to effect sizes obtained after less than six weeks of treatment using Enhanced Conversational Recast procedures (Plante et al. 2014 and in the current study). This is particularly striking because weekly intensity was nearly equivalent (120 minutes vs. 150 minutes per week), but total intervention intensity in terms of minutes was very different (~2880 minutes in Leonard et al., 2006, versus 750 minutes in the current study and Plante et al., 2014).

There are extreme differences in treatment duration between these examples. The long duration involved in Leonard et al. (2006) necessitates a discussion of caveats. In addition to being inefficient, long treatment durations may introduce the confounded variables associated with treatments that occur over long periods of time. As discussed in the introduction, particularly relevant in this particular case are maturation and the effect of outside treatment. Some children were receiving outside treatment, which could have contributed to gains on the target, control, or other linguistic forms. Signs of maturation were present in that children improved on control morphemes, and children in the control condition also demonstrated a significant change in morpheme production from baseline to six months later. This limits the extent to which effect sizes obtained in Leonard et al.’s study can be considered to be treatment-induced.

Outcome Measurement

A variety of outcome measures have been used across treatment studies. Some studies have included pre- and post-treatment standardized test data (e.g.,
Fey et al. 2003). Other studies have utilized measures such as MLU from spontaneous language samples (e.g., Yoder et al., 2011), in-treatment measures (e.g., Camarata et al., 1994), and generalization probes (e.g., Leonard et al., 2004; 2006). Fey & Finestack (2009) recommend that in early-stage research (feasibility & early efficacy studies), it is probably best to have measures closely tied to the target of the treatment (e.g., production of specific morphemes) and in later studies (e.g., effectiveness studies) more general measures of outcome might be warranted.

The choice of outcome measure can have a strong influence on the conclusions reached by a study. This is true for investigations of massed and spaced input in a therapeutic context (e.g., Smith-Lock et al, 2013b), as well as for studies involving conversational recast targeting morphosyntax in which distribution was not manipulated (e.g., Camarata et al., 1994; Leonard et al. 2004; 2006). As a case in point, the Smith-Locke et al (2013b) study considered scores on their Grammatical Elicitation Task. In this case, depending on which outcome measure and method of data analysis are selected, different conclusions are supported. Specifically, when gain scores are utilized, there is an apparent benefit to spacing. Conversely, when number of items correct on the grammar elicitation test at treatment end is used, there are no differences between the groups’ performances. In conclusion, outcome measures and methods of data analysis can meaningfully affect interpretation of results. This discrepancy offers a potential explanation for why Smith-Lock et al. reported an effect of distributed learning that was not evidenced in the current study and highlights the importance of considering the nature of the outcome measure or measures used.
Generalization Probes vs. In-treatment performance. In the present study, the generalization probes served as the primary outcome measure. Other studies of conversational recast studies have also used generalization probes as the primary measure of treatment effects (Leonard et al., 2004; 2006; 2008). In contrast, Camarata et al. (1994) utilized an in-treatment performance measure of spontaneous use of the target form. Data was collected during treatment sessions, which is different from collecting probe data in a testing context. Kamhi (2014) has recently argued that a distinction should be made between measures that reflect an immediate performance effect and outcome data that reflect true learning. For the present study, the probes collected during the treatment period were referred to as generalization probes because the term 'generalization' has typically been used to indicate that the target form is used in new, untrained contexts. Performance on generalization probes can be conceptualized as demonstrating learning of the target specifically, rather than generalization in a broader sense, say, to the use of additional rule-governed linguistic forms (i.e., grammatical morphology; Kamhi, 2014). Therefore, the treatment effect demonstrated by performance on generalization probes is a more appropriate measure of learning in an early efficacy study than an in-treatment measure that may simply be an immediate performance effect.

Rather than obtaining a single-point, end-treatment measurement, performance on generalization probes was tracked throughout the treatment period. This allows the trajectory or growth of learning or progress over the duration of the treatment program to be visualized and understood qualitatively. Of particular note
in this study was that in-treatment data closely tracked generalization data. Therefore, it was not the case that in-treatment performance had to reach a particular level before ‘generalization’ occurred. Instead, performance and generalization emerged in parallel, suggesting that the treatment procedure was directly responsible for promoting context-independent, generalized learning.

**Spontaneous Use.** Spontaneous use data provided an adjunct to generalization probes in terms of providing a measure of learning. Camarata et al. (1994) used time to first spontaneous production as a measure of treatment-induced growth. This metric was not selected because in a previous treatment study context (Plante et al., 2014), some children produced the target morpheme spontaneously in one treatment session but did not do so again in multiple subsequent treatment sessions. This is consistent with the notion that data collected within the context of a treatment session may be a reflection of immediate performance effects, rather than a measure of generalized learning. While caution is warranted for the interpretation of data gathered during treatment, it is the case that regular spontaneous use of morpho-syntax targets across the different contexts offered across days of treatment is meaningful relative to the purpose of intervention.

**Long-term retention.** It is important to evaluate retention of learning in order to provide additional support for the efficacy of treatment. If a child demonstrated learning (gains on the target or mastery) during the treatment period, but did not maintain that learning after the treatment had ended, the learning may have been
tenuous and vulnerable to forgetting (cf. Riches et al., 2005), or reflective of an immediate performance effect rather than true learning (cf Kamhi, 2014). On the other hand, if a child demonstrated learning during the treatment period and performed well on a follow-up measure collected after a period of time without treatment, evidence of retention of learning can support the efficacy of the treatment. For the present study, treatment occurred over a five-week period and follow-up occurred 7-11 weeks after the end of treatment. Therefore, in this case retention occurred over a period of time that was longer than that of the treatment period itself.

Factors Inherent to the Treatment that May Influence Outcomes

The ‘active ingredients’ of Conversational Recast that contribute to positive treatment outcomes are not fully understood. Although it is thought that naturalistic play contexts and contingency on child utterances makes these models of adult-like grammatical more effective than other treatments, these may not be the only or even the most important factors operating to make this treatment effective. Conversational recast involves environmental arrangement by the clinician to create communicative contexts in which the child is likely to attempt the target morpheme. Importantly, conversational recast therapy does not involve explicit instruction on grammaticality or a demand for imitation from the child, which may reduce the burden on the child. Furthermore, conversational recast provided in a linguistically rich environment, one in which the needed vocabulary is supplied for the child and utterances are encouraged, but not demanded, through elicitation and prompts. In
combination, these elements of conversational recast all plausibly contribute to the efficacy of the basic treatment technique.

It is thought that Enhanced Conversational Recast confers benefit above and beyond traditional conversational recast because of the incorporation of high variability input and attentional cueing. Following from learning theory (Alt et al., Plante et al., 2014) providing linguistic variability in terms of providing a large variety of the elements of input that are free to vary (such as the verb stem), can help the learner focus on stable elements of the input (the grammatical inflection that is the target). Additionally, the variety of play contexts in which the recasts are provided create many different engaging and semantically-supported opportunities to hear the target form.

While Enhanced Conversational Recasts are delivered in natural play contexts, with characteristics such as natural prosody and timing that make recasts conversation-like, the clinician must not only provide appropriately delivered recasts, but he or she must ensure that the child receives (hears and attends to) the recast. In terms of treatment dosage, the difference between a dose administered (an appropriate conversational recast) and a dose received may be a critical difference in learning outcomes for conversational recast treatment for grammatical morpheme errors. The need to have the child attend to the conversational recast may make the interaction less conversation-like, but it may be a critical difference for children with SLI, who are reported to have attentional deficits (Spaulding, Plante & Vance, 2008), and whose language development has not benefitted sufficiently from natural conversational input. Also, treatment in this study and the
previous (Plante et al., 2014) has been provided in a one-on-one context, which may facilitate the child’s attending to the clinician, and allows the clinician to focus on a single individual during treatment.

There is also a role of dosage and treatment intensity in the design and planning of effective treatment. It is well-established that the dose number of conversational recasts needs to be sufficient to effect change. Recasts are common in the speech of adults to children outside of a therapeutic context, and almost all the language a child hears is a potential model of adult-like morphosyntax. We know that children with SLI need more exposures to learn linguistic structures than do typically-developing children (e.g., Oetting, Rice, & Swank, 1995; Rice, Oetting, Marquis, Bode, & Pae, 1994). Children with SLI who are delayed in acquisition of grammatical morphemes have not been able to learn the morphosyntax present in their linguistic environment through natural amounts of exposure.

Factors Inherent to the Child that May Influence Outcomes

Children in this treatment study achieved a very wide range of effect sizes for treatment. The majority of children demonstrated evidence of an excellent or good treatment response, but some had a poor response to treatment. Children were classified by treatment response for descriptive purposes. Six children (three in the spaced condition and three in the massed condition) were considered ‘excellent’ responders to the treatment in this study due to achieving the highest effect sizes ($d = 15.01$ or greater) and achieving and maintaining ceiling performance during the treatment period. Five other children were characterized as ‘good’ responders
based on treatment effect sizes of $d = 2.18-8.0$. Another five children were considered relatively ‘poor’ responders. Effect sizes for poor responders were $d = 1.01$ and smaller.

While all the factors that may contribute to differential responses to treatment are not known, some existing treatment literature addresses response variability. The use of correlational analyses in this study also provided some data that pertains to child factors that may influence treatment outcomes.

Yoder, Molfese, and Gardner (2011) reported that children’s success in treatment was related to initial language levels. The small and non-significant relationship between pretreatment performance in the baseline period and treatment effect size for children in the present study ($r = .093$) does not provide converging evidence for Yoder and colleagues’ claim. Additionally, SPELT-P2 scores, an index of overall language severity, were not correlated with treatment effect sizes. Therefore, a child’s language level prior to treatment, either specific to the grammatical form to be targeted for treatment (pretreatment use), or more broadly construed (SPELT-P2), did not relate to treatment outcomes. It is possible that pretreatment performance is correlated with outcomes when a treatment is not efficacious overall, and this appeared to be the case in the Yoder et al. (2011) study. In contrast, the Enhanced Conversational Recast treatment utilized in the present study is efficacious. In the present study, even children with very low or zero percent accuracy pretreatment were able to make significant gains, and ceiling performance was reached for some.
Correlational analyses indicate that the measure of nonverbal IQ (K-ABC) obtained prior to the start of the study was positively and significantly correlated with treatment effect size ($r = .599$). It is also worth noting that none of the subjects had nonverbal IQs in a range consistent with intellectual disability. Indeed, the lowest nonverbal measure was 84, just below one standard deviation from the population mean. Although non-verbal IQ was significantly positively correlated with treatment effect size, this relationship should be interpreted with caution. There was a limited range of non-verbal IQ, with particularly limited representation of the lower end of non-verbal IQ scores that are above the cut-off for intellectual disability. Also, this correlated relationship does not imply that children with lower non-verbal IQ scores do not or cannot respond to this treatment. Indeed, Plante et al. (2014) involved two children with nonverbal IQ scores lower than 84 that responded well to Enhanced Conversational Recast treatment.

Additionally, a significant positive correlation ($r=.490$) was found between a measure of receptive vocabulary (PPVT-4) and treatment effect size. This relationship indicates that those children with small lexicons may struggle to learn with Enhanced Conversational Recast techniques. However, there are exceptions to this trend. For example, subject 502, whose PPVT-4 score was among the lowest, still showed an excellent response to treatment. Nevertheless, caution is warranted in applying this particular treatment to groups of children with low vocabulary knowledge. For example, younger children, such as 3-year-olds with SLI, may have vocabulary limitations that mitigate the efficacy of this treatment approach.
Likewise English language learners may have insufficient vocabulary to benefit from this particular treatment.

While this study was not designed to identify and analyze factors of the child that may influence intervention outcomes, it is necessary to point out the potential role of these factors. Children with SLI differ in their response to treatments. This population also represents a range of language and cognitive abilities that may contribute to treatment outcomes.

The long-term retention measure provides evidence for the stability of learning demonstrated during the treatment period. Although follow-up occurred through structured generalization probe procedures, the long duration of seven to eleven weeks from the end of treatment allows performance to be characterized as long-term retention. In total, nine children of 13 who returned for follow-up demonstrated retention of learning.

It is a reasonable possibility that some of the children in the ‘poor responder’ category would have demonstrated learning of the target morpheme had the treatment period been extended. For example, one child (305) appeared to make gains during the final days of the treatment program. Her end treatment performance average was 27%, but she achieved 50% accuracy on the follow-up measure. Her retention may have been even better had she had more treatment days during which to stabilize her performance. In contrast, children in the Excellent and Good responder groups tended to achieved mastery of their morpheme targets earlier in treatment. This general pattern suggests that some children required more treatment sessions than other children to demonstrate
learning of the target morpheme. However, it is often the case in treatment studies that some participants do not demonstrate a treatment effect, even when treatments are continued over long durations (e.g., Leonard et al., 2006). This can be true even when non-responding children have seemingly similar language deficits and profiles as those who do respond. This was true of the present study as well. It is possible that these children, often termed non-responders, require a different type of treatment. One goal of treatment research involves not only efforts to increase overall treatment efficacy, but also make treatments efficacious for more children. Over time, with evidence from carefully designed and reported treatment studies, it may be possible to identify profiles of children who might respond more or less readily to Enhanced Conversational Recast treatment.

Limitations

This experimental study was carefully designed to have good experimental control over a number of factors that have been left free to vary in previous studies. However, given the nature of treatment research, particularly treatment involving young children, it is not possible to control for and balance across conditions for all experimental variables. Due to the limited sample size, and variety of linguistic needs of individual children (i.e., the grammatical forms the child was not using in the baseline period), it was not possible to balance morpheme perfectly across conditions. Treatment and control targets were balanced across children and conditions to the extent possible. This minimized, but did not completely eliminate, the chance that a difference in treatment effects across morphemes assigned to children and across groups were due to differences in the difficulty of acquisition of
any one morpheme over the other. There was a fair amount of balance of morpheme targets across conditions, allowing for replication in both conditions for all but one of the morphemes (“doesn’t” was treated for only one child and a control morpheme for one other child, but not replicated across children). There was good balance between grammatical forms that served as treatment targets and controls (all morphemes treated were also controls). This reduced the likelihood that the specific morphemes targeted for treatment were more likely to change than those designated as control morphemes.

That said, there appeared to be a general relationship between the degree of response to treatment and the category of grammatical form that served as the treatment target. All children with free morpheme targets achieved higher effect sizes than all children with verb morpheme targets. This study provided a higher proportion of children who were trained on free morphemes than in the previous Enhanced Conversational Recast study (Plante et al., 2014). Only one child was treated for a free morpheme error (the pronouns “she”) in that study.

Not all elements of dosage that could contribute to the strength of treatment effect could be investigated in this single study. For example, some authors (e.g., Proctor-Williams & Fey, 2007; Yoder, Fey & Warren, 2012) have suggested that the role of dose rate (total recasts per unit of time) and dose distribution (spacing between each individual recast within a treatment period) may be important variables for input-based treatments. In the present study, dose rate was held constant across the two treatment conditions and the actual distribution of recasts during the treatment session was not monitored beyond the constraint of delivering
8 recasts in each 10-minute period of treatment. Therefore, no conclusions can be drawn about the role of rate and distribution from the present study, beyond the conclusion that the rate of input of 8 recasts in 10 minutes (one recast per 1.25 minutes, or .8 recasts per minute) can lead to significant treatment effects.

This study involved a group of 4- and 5-year-old children with SLI, and provided a replication of the robust effects obtained through providing Enhanced Conversational Recast. In addition to replication of treatment efficacy, this study adds to the evidence base by reporting the very low and non-significant effect size associated with massed and spaced group differences. Given these two findings, it would be reasonable to expect that other children this age with SLI would respond in a similar way as the children in this study to Enhanced Conversational Recast, delivered in either massed or spaced doses.

As an early efficacy study, the purpose of this study was to demonstrate "a relationship between an intervention and an outcome." under ideal or near ideal conditions (Fey & Finestack, 2009). The next stage of treatment research that can build upon these findings is later efficacy studies. Strong later efficacy studies may compare the new treatment to a standard or existing treatment, under more generalizable conditions. These studies may also be designed to answer additional, lingering questions from previous research studies. For example, the follow-up measure utilized in this study did not include a language sample as an outcome measure, which provides a more general measure of language gains than either generalization probes or spontaneous use as measured here. A later efficacy study designed to further refine Enhanced Conversational Recast as a treatment, or to
identify additional parameters related to treatment dosage, could include a spontaneous use measure. This could be used as a measure of generalized language benefit, beyond the morphemes treated, or used after a period of no treatment following the end of a treatment program.

Conclusion

This study demonstrated a method to investigate a particular component of an already established evidence-based treatment program for children SLI. In summary, on average there was a significant (positive) and equivalent treatment effect for children in both conditions, but no improvement on control morphemes, demonstrating the efficacy of the intervention. The outcomes of this study for children in both conditions replicate previous findings supporting the effectiveness of Enhanced Conversational Recast provided daily for five weeks for children with SLI (Plante et al., 2014). There was no effect of spaced or massed dose schedule within a treatment day, or over the treatment period. This provides flexibility for clinicians to schedule treatment sessions either massed or distributed throughout the day for children with specific language impairment.

The results of this study can immediately support the treatment choices of speech-language pathologists who provide services to children with SLI. Furthermore, the findings of this study may inform other treatment researchers who are interested in the role of dose schedule in intervention programs, and promote the investigation of dose schedule in other clinical contexts and across different populations.
Appendix A: Examples of obligated and non-obligated contexts for morpheme use.

To elicit “she” or the use of 3rd person singular “-s”

**Obligated Contexts**

<table>
<thead>
<tr>
<th>Obligated Context</th>
<th>Not Obligated</th>
</tr>
</thead>
<tbody>
<tr>
<td>What happens with the girl?</td>
<td>What does she do?</td>
</tr>
</tbody>
</table>
| or
| Tell me what happens with Sara. | |

(Target response: **She** verbs) (Possible response: Unmarked verb, no pronoun)

To elicit 3rd person singular “-s”

**Obligated Context**

These guys flip. Tell me what happens with that guy.

(Target response: **He** flips) (Possible response: Unmarked verb)

To elicit auxiliary "is verb-ing”

**Obligated Context**

Watch her spin [demonstrate action continuously]. What’s happening?

(Target response: The girl is spinning) (Possible response: Spinning)

To elicit regular past tense “-ed”

**Obligated Contexts**

<table>
<thead>
<tr>
<th>Obligated Context</th>
<th>Not Obligated</th>
</tr>
</thead>
<tbody>
<tr>
<td>I saw him bump the wall. What happened?</td>
<td>What did he do?</td>
</tr>
</tbody>
</table>

(Target response: He bumped the wall) (Possible response: Bump the wall.)

He was jumping, but now he is all done. Tell me what the horse did.

(Target Response: The horse jumped) (Possible response: Jumping.)
References


