

Title: Maximizing outcomes for preschoolers with developmental language disorders:

Testing the effects of a sequentially targeted naturalistic intervention

National Clinical Trial (NCT) Identified Number: NCT03782493

Principal Investigators: Pamela Hadley, Ann Kaiser, Megan Roberts

Funded by: NIDCD

Version Date: 11/13/20

IRB Modification Number: 45

Table of Contents

STATEMENT OF COMPLIANCE	1
1 PROTOCOL SUMMARY	1
1.1 Synopsis	1
1.2 Schema	3
1.3 Schedule of Activities (SoA)	4
2 INTRODUCTION	6
2.1 Study Rationale	6
2.2 Background	8
2.3 Risk/Benefit Assessment	10
2.3.1 Known Potential Risks	10
2.3.2 Known Potential Benefits	11
2.3.3 Assessment of Potential Risks and Benefits	11
3 OBJECTIVES AND ENDPOINTS	11
4 STUDY DESIGN	13
4.1 Overall Design	13
4.2 Scientific Rationale for Study Design	14
4.3 Justification for Dose	15
4.4 End of Study Definition	16
4.5 Inclusion Criteria	16
4.6 Exclusion Criteria	16
4.7 Lifestyle Considerations	16
4.8 Screen Failures	17
4.9 Strategies for Recruitment and Retention	17
5 STUDY INTERVENTION	19
5.1 Study Intervention(s) Administration	19
5.1.1 Study Intervention Description	19
5.1.2 Dosing and Administration	20
5.2 Preparation/Handling/Storage/Accountability	27
5.2.1 Acquisition and accountability	28
5.2.2 Formulation, Appearance, Packaging, and Labeling	28
5.2.3 Product Storage and Stability	28
5.2.4 Preparation	28
5.3 Measures to Minimize Bias: Randomization and Blinding	28
5.4 Study Intervention Compliance	29
5.4.1 Rescue Medicine	29
6 STUDY INTERVENTION DISCONTINUATION AND PARTICIPANT DISCONTINUATION/WITHDRAWAL	29
6.1 Discontinuation of Study Intervention	29
6.2 Participant Discontinuation/Withdrawal from the Study	29
6.3 Lost to Follow-Up	30
7 STUDY ASSESSMENTS AND PROCEDURES	30
7.1 Efficacy Assessments	30
7.2 Safety and Other Assessments	32
7.3 Adverse Events and Serious Adverse Events	35
7.3.1 Definition of Adverse Events (AE)	35
7.3.2 Definition of Serious Adverse Events (SAE)	35

7.3.3	Classification of an Adverse Event.....	36
7.3.4	Time Period and Frequency for Event Assessment and Follow-Up	36
7.3.5	Adverse Event Reporting	37
7.3.6	Serious Adverse Event Reporting.....	37
7.4	Unanticipated Problems.....	37
7.4.1	Definition of Unanticipated Problems (UP).....	37
7.4.2	Unanticipated Problem Reporting.....	38
7.4.3	Reporting Unanticipated Problems to Participants	38
8	STATISTICAL CONSIDERATIONS.....	38
8.1	Sample Size Determination	38
8.2	Populations for Analyses	39
8.3	Statistical Analyses	39
8.3.1	Safety Analyses	46
8.3.2	Baseline Descriptive Statistics	46
8.3.3	Planned Interim Analyses	46
8.3.4	Sub-Group Analyses.....	46
8.3.5	Tabulation of Individual participant Data	47
9	SUPPORTING DOCUMENTATION AND OPERATIONAL CONSIDERATIONS	47
9.1	Regulatory, Ethical, and Study Oversight Considerations.....	47
9.1.1	Informed Consent Process	47
9.1.2	Study Discontinuation and Closure	48
9.1.3	Confidentiality and Privacy	48
9.1.4	Future Use of Stored Specimens and Data	49
9.1.5	Key Roles and Study Governance	49
9.1.6	Safety Oversight.....	50
9.1.7	Clinical Monitoring.....	51
9.1.8	Quality Assurance and Quality Control.....	51
9.1.9	Data Handling and Record Keeping.....	52
9.1.10	Protocol Deviations	53
9.1.11	Publication and Data Sharing Policy	54
9.1.12	Conflict of Interest Policy.....	55
9.2	Additional Considerations	55
9.3	Abbreviations	55
9.4	Protocol Amendment History	56
10	REFERENCES.....	58

STATEMENT OF COMPLIANCE

The trial will be carried out in accordance with International Conference on Harmonisation Good Clinical Practice (ICH GCP) and the following:

- United States (US) Code of Federal Regulations (CFR) applicable to clinical studies (45 CFR Part 46, 21 CFR Part 50, 21 CFR Part 56, 21 CFR Part 312, and/or 21 CFR Part 812)

National Institutes of Health (NIH)-funded investigators and clinical trial site staff who are responsible for the conduct, management, or oversight of NIH-funded clinical trials have completed Human Subjects Protection and ICH GCP Training.

The protocol, informed consent form(s), recruitment materials, and all participant materials will be submitted to the Institutional Review Board (IRB) for review and approval. Approval of both the protocol and the consent form must be obtained before any participant is enrolled. Any amendment to the protocol will require review and approval by the IRB before the changes are implemented to the study. In addition, all changes to the consent form will be IRB-approved, and a determination will be made regarding whether a new consent needs to be obtained from participants who provided consent, using a previously approved consent form.

1 PROTOCOL SUMMARY

1.1 SYNOPSIS

- Title:** Maximizing outcomes for preschoolers with developmental language disorder: Testing the effects of a sequentially targeted naturalistic intervention.
- Study Description:** The objective of the proposed study is to evaluate the efficacy of the Enhanced Milieu Teaching-Sentence Focus (EMT-SF) intervention, implemented by caregivers and interventionists, relative to a control condition enrolling 108 30-month-old children and their caregivers. The central hypothesis is that intervention will result in better overall child language skills at 49 months of age.
- Objectives:** To evaluate the effects of the Enhanced Milieu Teaching-Sentence Focus (EMT-SF) intervention in comparison to a business-as-usual control condition.
- Primary Objectives
- 1.1 Compare overall language ability at 49 months of age.
 - 1.2 Compare growth in vocabulary skills during intervention from 30 to 49 months of age.
 - 1.3 Compare growth in grammar skills during intervention from 30 to 49 months of age.

Secondary Objectives

- 2.1 Compare differences in short-term vocabulary outcomes at 36 months of age.
- 2.2 Compare differences in short-term grammar outcomes at 42 months of age.
- 2.3 Compare differences in child communication and interaction in real world settings at 49 months of age.
- 2.4 Compare rates of developmental language disorder diagnoses at 49 months of age.
- 2.5 Compare differences in decontextualized language use at 49 months of age.
- 2.6 Compare differences on narrative skills at 49 months of age.
- 2.7 Compare differences in internalizing behaviors at 49 months of age.

Exploratory Objectives

- 3.1 Examine the extent to which baseline receptive language skills moderate intervention outcomes.
- 3.2 Examine the extent to which biological sex moderates intervention outcomes.
- 3.3 Examine the extent to which family history of language impairment moderates intervention outcomes.
- 3.4 Examine the extent to which disruptive behaviors at baseline moderate intervention outcomes.
- 3.5 Examine the extent to which caregiver use of child-directed, developmentally appropriate target language mediates intervention outcomes.
- 3.6 Examine the extent to which child growth in vocabulary and grammar during intervention mediates intervention outcomes.

Endpoints:

Primary Endpoint:

- 1.1 CELF-P2 at 49 months of age.
- 1.2 Total number of different spoken word (NDW) roots from 30 to 49 months of age.
- 1.3 Latent variables (verbs, subjects, tense/agreement, productive syntax, clauses per utterance) from 30 to 49 months of age.

Secondary Endpoints:

- 36 months of age (6 months after study start, 2.1)
- 42 months of age (9 months after study start 2.2)

- 49 months of age (18 months after study start; 2.3, 2.4, 2.5, 2.6, 2.7, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6)

Sample Size 108

Study Population: 30-month-old children (between 30-31 months of age) with receptive-expressive language delays within a 3 hour drive from the university sites in Nashville, TN, Chicago IL or Champaign, IL. .

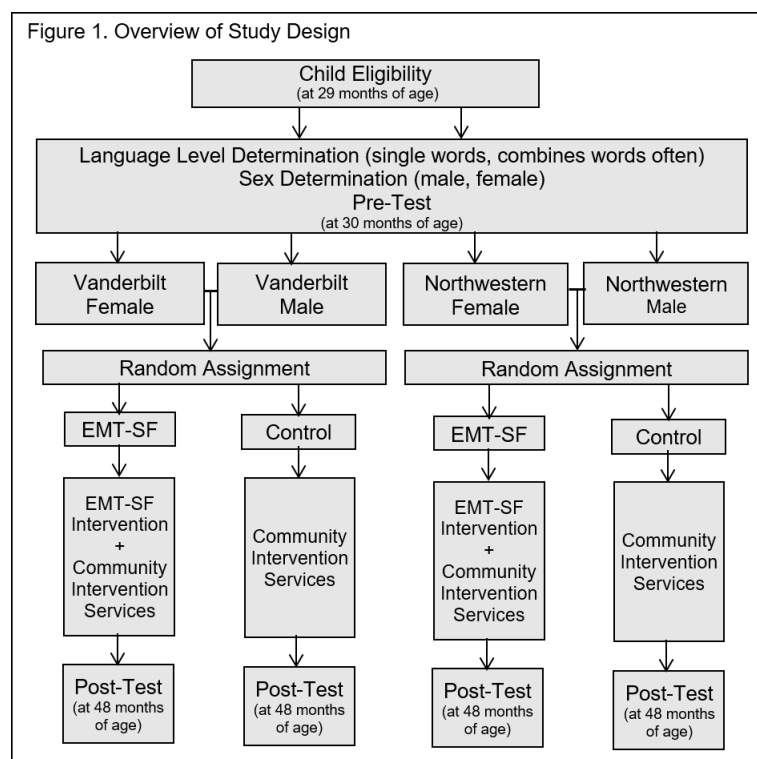
Phase: 2

Description of Sites/Facilities Enrolling Participants: The project includes two primary locations for enrolling participants, including sites situated in the Department of Communication Sciences and Disorders (CSD) at Northwestern University and in the Department of Special Education at Peabody College at Vanderbilt University and the Vanderbilt Kennedy Center for Research on Human Development.

Description of Study Intervention: The study intervention is a behavioral intervention which will include individually teaching caregivers to use the intervention strategies from the Enhanced Milieu Teaching-Sentence Focus (EMT-SF) intervention using a manualized protocol (Teach-Model-Coach-Review). Caregivers will participate in 66 virtual intervention sessions across 18 months, targeting vocabulary and grammar as well as the transition to decontextualized language.

Participant Duration: 18-19 months (depending on age of enrollment)

1.2 SCHEMA



1.3 SCHEDULE OF ACTIVITIES (SOA)

Table 1. Schedule of Activities

	Recruitment Day -90 to -15	Screening, Virtual Day -14 to -1	Enrollment/Baseline, Virtual 30-31 ^a months of age Day 1	Treatment sessions 1-24	33 months of age, Virtual Day 90 (+14 days)	Treatment sessions 25-36	36 months of age, Virtual Day 180 (+14 days)	Treatment sessions 37-48	39 months of age, Virtual Day 270 (+14 days)	Treatment sessions 49-60	42 months of age, Virtual Day 360 (+14 days)	Treatment sessions 61-63	45 months of age, Virtual Day 450 (+14 days)	Treatment sessions 64-66	49 months of age, Virtual Day 540 (+7 days)	49 months of age, Virtual Day 547 (+7d days)
Procedures																
Recruitment	P1 ^b															
Study consent			P1													
Randomization			C ^c													
Adverse event monitoring		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
Evaluations																
Demographics			P1				P1				P1				P1	
Community services form			P1		P1		P1		P1		P1		P1		P1	
Autism Screening Survey		P1														
Macarthur-Bates Communication and Development Inventories: Words and Sentences (MCDI-WS)		P1			P1		P1									

	Recruitment Day -90 to -15	Screening, Virtual Day -14 to -1	Enrollment/Baseline, Virtual 30-31 ^a months of age Day 1	Treatment sessions 1-24	33 months of age, Virtual Day 90 (+14 days)	Treatment sessions 25-36	36 months of age, Virtual Day 180 (+ 14 days)	Treatment sessions 37-48	39 months of age, Virtual Day 270 (+14 days)	Treatment sessions 49-60	42 months of age, Virtual Day 360 (+14 days)	Treatment sessions 61-63	45 months of age, Virtual Day 450 (+14 days)	Treatment sessions 64-66	49 months of age, Virtual Day 540 (+7 days)	49 months of age, Virtual Day 547 (+ 7d days)
Focus on the Outcomes of Children Under Six (FOCUS)			P1		P1		P1		P1		P1		P1		P1	
Multiple Assessment of Preschool Disruptive Behavior (MAP-DB)			P1		P1		P1		P1		P1		P1		P1	
Mullen Scales of Early Learning – Visual Reception subtest (MSEL VR)		C														
Preschool Language Scales – 5 th Edition (PLS-5)		C														
Telemedicine-based ASD Evaluation Tool for Toddlers and Young Children (TELE-ASD-PEDS)		C														
Autism Diagnostic Interview – Revised (ADI-R)		P														
Language Sample (LS)			C		C		C		C		C		C		C	
Caregiver-Child Interaction (CCX)			C		C		C		C		C		C			C
Peabody Picture Vocabulary Test – 5 th Edition (PPVT-5)							C								C P1	
Expressive Vocabulary Test – 3 rd edition (EVT-3)							C								C	
Structured Photographic Expressive Language Test – Preschool 2 nd edition (SPELT-P2)											C				C	
Test of Early Grammatical Impairment - 3 rd person singular & past tense probes (TEGI)											C				C	
Clinical Evaluation of Language Fundamentals Preschool – 2 nd edition (CELF-P2)																C
Renfrew Bus Story – North American Edition (RBS-NA)															C	
Goldman-Fristoe Test of Articulation – 3 rd edition (GFTA-3)																C
Child Behavior Checklist (CBCL) 1 ½-5			P1												P1	
Parent Self-Efficacy and Parent Knowledge of Language Survey			P1				P1				P1				P1	
Printed Intervention Materials & Toy Distribution (Both Intervention & Control Group)																
Printed intervention materials and toys distributed to caregivers			P1		P1		P1		P1		P1		P1			
Intervention and Progress Monitoring (Intervention Group only)																
EMT-SF Intervention: Core Vocabulary (Phase 1)				C P1		C P1		C P1								

	Recruitment Day -90 to -15	Screening, Virtual Day -14 to -1	Enrollment/Baseline, Virtual 30-31 ^a months of age Day 1	Treatment sessions 1-24	33 months of age, Virtual Day 90 (+14 days)	Treatment sessions 25-36	36 months of age, Virtual Day 180 (+ 14 days)	Treatment sessions 37-48	39 months of age, Virtual Day 270 (+14 days)	Treatment sessions 49-60	42 months of age, Virtual Day 360 (+14 days)	Treatment sessions 61-63	45 months of age, Virtual Day 450 (+14 days)	Treatment sessions 64-66	49 months of age, Virtual Day 540 (+7 days)	49 months of age, Virtual Day 547 (+ 7d days)
EMT-SF Intervention: Sentences (Phase 2)										C P1		C ^d P1		C ^d P1		
EMT-SF Intervention: Decontextualized Language (Phase 3)												C P1		C P1		
6-minute unsupported caregiver- child interaction prior to intervention session				C		C		C		C		C				
Video recording - Fidelity check				P1		P1		P1		P1		P1				

^a Timepoints are approximate and may vary up to two months based on age of entry and assessment window

^b P1: Caregiver 1, participating in assessments and intervention (if applicable)

^c C: Child participating in the study

^d Phases 1 & 2 may continue up to 49 months of age depending on progress on diverse sentences.

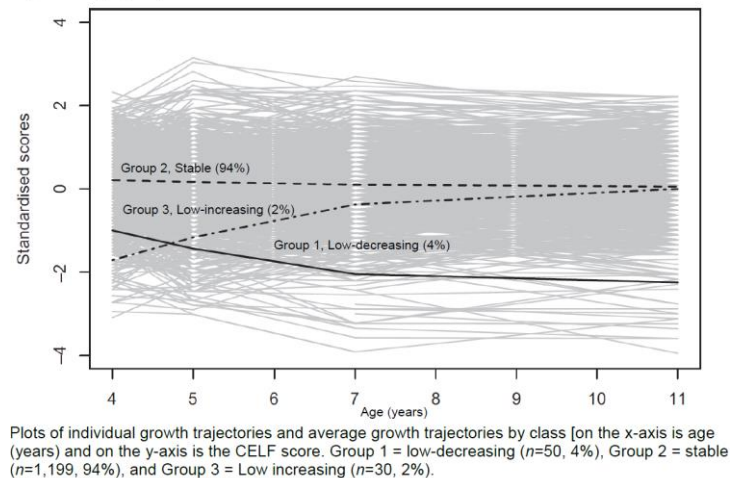
2 INTRODUCTION

2.1 STUDY RATIONALE

Between 7-10% of children have developmental language disorders (DLD) without other developmental concerns at kindergarten entry (Tomblin et al., 1997; Rice, Tomblin, Hoffman, Richman, & Marquis, 2004). Children with DLD have extreme difficulty understanding and producing language, relative to age expectations, in the absence of an associated biomedical condition, intellectual disability, or global developmental delay (Bishop, Snowling, Thompson, & Greenhalgh, 2017). Clinically significant language disorders, as defined by the DSM-V, are characterized by reduced vocabulary, limited sentence structure, and impairments in discourse (e.g., conversation, storytelling, explanations) that affect every day functioning. Clinically significant language disorders are identified by performance that falls below the 10th percentile on two or more standardized measures distributed across language comprehension, production, vocabulary, grammar, and discourse (i.e., Tomblin et al., 1997). Children with DLD may also have low nonverbal intelligence quotients (i.e., NVIQ > 70; Bishop et al., 2017). In the research literature, studies have focused primarily on children with specific language impairment (SLI), the most common subtype of DLD. Children with SLI have clinically significant language disorders with NVIQ greater 85. However, regardless of NVIQ, clinically significant language disorders occurring at age 5 are likely to persist.

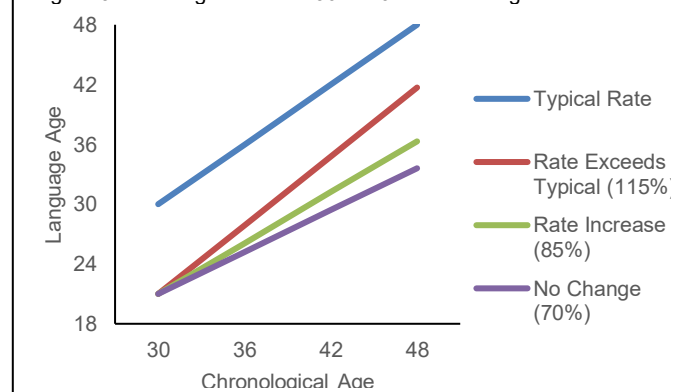
Although spontaneous recovery may occur between 2 and 4 years of age (Reilly, Wake, & Ukoumunne, 2010), clinically significant language disorders at age 4 place children at significant risk for poor academic, behavioral, and social outcomes across the school years and into adulthood (Walker, Greenwood, Hart, & Carta, 1994; Whitehouse, Watt, Line, & Bishop, 2009; Clegg, Hollis, Mawhood, & Rutter, 2005; Johnson, Beitman, & Brownlie, 2010; Yew & O’Kearney, 2013). Longitudinal studies from the US, UK, and Australia indicate that language development trajectories are remarkably stable from age 4 through adolescence across all levels of language ability. For example, McKean et al. (2017) demonstrated that 94% of children showed stable growth trajectories on a standardized measure of language development between ages 4 and 11. (See Figure 2.) Thus, a profile of DLD, present at age 4, is highly likely to persist (McKean et al., 2017; Norbury & Sonuga-Barke, 2017; Conti-Ramsden, Clair, Pickles, & Durkin, 2012; Rice & Hoffman, 2015; Tomblin, Zhang, Buckwalter, & O’Brien, 2003). Therefore, opportunities to shift children’s language trajectories are essential prior to age 4.

Figure 2. Language trajectories from 4 to 11.



Brief naturalistic interventions such as Enhanced Milieu Teaching (EMT) can significantly increase expressive vocabulary and reduce behavior problems in toddlers with receptive and expressive language delays in comparison to no treatment groups (Roberts & Kaiser, 2015; Curtis, Kaiser, Estabrook, & Roberts, 2018). However, brief EMT intervention alone does not appear to alter language trajectories sufficiently to reduce the developmental gaps in vocabulary or syntax between young children with both receptive and expressive language delays and their typically developing peers (Hampton, Kaiser, & Roberts, 2017; Hadley, Kaiser, & Roberts, 2018). For example, 30-month-old children with clinically significant language delays (e.g., 30% delay, 21-month language age at a chronological age of 30 months) are learning language at approximately 70% of the average rate of typical peers. Although brief interventions may increase their learning rates (e.g., 85%) relative to pre-intervention rates (e.g., 70%), children with language delays will fall further behind same-age peers on standardized measures unless their learning rates exceed that of typical peers (e.g., 115%) for a sustained period of time (see Figure 3). Thus, to reduce developmental gaps on standardized language measures, interventions must accelerate the rate of development. To do this, clinical research must move beyond brief interventions on single, early dimensions of language (e.g., vocabulary, grammatical morphemes), target critical transitions between dimensions, and provide explicit support for grammatical development. No studies to date have evaluated the potential for sustained early intervention to maximize language outcomes by

Figure 3. Learning rates from 30 to 49 months of age



using a strategic sequence of treatment targets to facilitate the transition from vocabulary to syntax and the use of language in contextualized to decontextualized situations.

The proposed intervention is designed to accelerate language growth and sustain developmental momentum in order to close gaps between children at risk for DLD and their typically developing peers by targeting key transitions in language development and by expanding the child-only language intervention to include parent education and coaching during the preschool years. This intervention has the potential to transform preschool service delivery models, to strengthen the spoken language foundation needed for school success, and prevent or reduce the long-term risks of DLD.

2.2 BACKGROUND

Children with DLD have substantial difficulties producing and/or understanding language; these difficulties affect everyday activities and exist in the absence of a biomedical condition (Bishop et al., 2017). Approximately 7% to 10% of all children at school entry are affected by DLD (Tomblin et al., 1997; Rice et al., 2004). Children with DLD typically experience late onset of first words and word combinations and difficulties mastering grammar that become evident over time (Leonard, 2014; Rice, 2013). DLD has negative consequences on social interactions, peer relationships, behavior, literacy, educational progress, and mental health outcomes (Clegg et al., 2005; Johnson et al., 2010; Yew & O'Kearney, 2013; Whitehouse et al., 2009), yet DLD often remains unidentified and untreated in early childhood (Tomblin et al., 1991; Norbury & Sonuga-Barke, 2017).

Toddlers at-risk for DLD are unlikely to receive early intervention, despite the greater potential to maximize language growth by starting intervention early. Law and colleagues (Law, Charlton, & Asmussen, 2017) have called for a public health approach to address the needs of toddlers who show substantial early language delays. In a tiered prevention/ intervention model, targeted interventions identify and intervene with individuals or groups of children whose risk of developing specific problems (e.g., DLD) is high. Developmental surveillance or 'watchful waiting' is the prevailing policy guiding US early intervention for toddlers with language delays and no other established biomedical conditions or socioeconomic risk. This policy is based on findings that some early delays resolve spontaneously without direct intervention (Dollaghan, 2013; Paul, 1996; Paul, Bishop, & Leonard, 2000). However, evidence for spontaneous recovery varies by language delay profile and by aspect of the language system. Early language delays are estimated to resolve for 60% of toddlers with expressive only (E) delays, but for only 25% of children with receptive *and* expressive (RE) delays (Law, Boyle, Harris, Harkness, & Nye, 2000). Additionally, recovery estimates are often based on vocabulary and global receptive and expressive language measures, not on assessment of grammatical development, the most vulnerable of aspect of language associated with DLD (Rescorla, Roberts, & Dahlsgaard, 1997; Rice, Taylor, & Zubrick, 2008). In the absence of intervention, developmental gaps between children at risk for DLD and their typically developing peers may increase because the processes of language learning are slowed or disrupted. For example, compared to children with typical language development, young children with significant production and comprehension delays focus on learning early vocabulary for an extended period. This slower vocabulary growth from 24 to 30 months leads to substantial grammatical delays at 36 months (i.e., more than 3 *SD* below the mean for typical children (Rescorla, Mirak, & Singh, 2000). Thus, for children with RE, the early processes for rapid language development are disrupted, and the risk for DLD becomes increasingly apparent as grammar emerges.

In a meta-analysis, Fisher (2017) examined predictors of expressive language outcomes between 3 and 5 years in 20 studies with 2,134 late-talking toddlers under 30 months. Three significant risk factors were identified: (a) expressive vocabulary size, (b) receptive language, and (c) socioeconomic status. Family history was not a significant predictor, but it was examined in only 6 studies. In a second meta-analysis, Rudolph (2017) used studies of children with a diagnosis of specific language impairment (SLI) after age 4, and examined the predictive value of case history variables for 6,275 children with SLI and 206,709 typically developing children. Four case history variables emerged as clinically significant: (a) maternal education level, (b) 5-min Apgar score, (c) birth order, and (d) biological sex. Family history and prematurity also were statistically significant, but did not reach clinical significance. McKean et al. (2017) identified low family literacy and low birth weight as significant predictors of children whose language problems increased in severity after age 4. Based on this empirical evidence, we propose that children with very small expressive vocabularies with receptive language delays at 30 months are at high risk for DLD, and that male sex, lower socioeconomic status, and a positive family history of language or learning disability increase risk for DLD.

Brief targeted early intervention can improve vocabulary outcomes in toddlers with receptive and expressive delay but may not close the developmental gap. Short-term, caregiver-implemented language interventions have been shown to accelerate vocabulary growth for toddlers with receptive and expressive (RE) delays (Roberts & Kaiser, 2015; Curtis et al., 2017; Roberts & Kaiser, 2011; Roberts & Kaiser, 2012). However, short-term intervention does not appear to be sufficient to reduce developmental gaps between toddlers with RE delays and typically developing peers (Hampton, Kaiser, & Roberts, 2017). Intensive language intervention for toddlers with RE delays may be necessary to establish a positive developmental trajectory leading to more optimal language development. Establishing a positive, sustained developmental trajectory for these children requires intervention that does more than improve expressive vocabulary. Early intervention must harness developmental momentum by: (1) building language comprehension; (2) supporting rapid word learning; (3) facilitating the transition from single words to syntax; and (4) teaching caregivers, as critical communication partners, to use language support strategies and to modify language input in response to children's changing developmental needs.

We propose to study the effects of an intensive, caregiver-implemented early language intervention specifically designed for toddlers at risk for DLD: Enhanced Milieu Teaching-Sentence Focus (EMT-SF). EMT-SF combines several empirically validated, active ingredients that provide high-quality, developmentally appropriate language input. EMT-SF builds on evidence from our early caregiver-implemented vocabulary-focused intervention for children with RE delays (Roberts & Kaiser, 2015), adds a theoretically-motivated sequence of treatment targets guided by linguistic theories of event structure in simple sentences (Bowers, 1993; Pustejovsky, 1991; Ramchand, 2008), and extends caregiver training to address targets across changing developmental contexts. The EMT-SF intervention is designed to promote flexible use of vocabulary in diverse, simple sentences as a catalyst for syntactic growth (Hadley, Rispoli, & Holt, 2017a; Hadley et al., 2017b). EMT-SF combines an effective process (Teach-Model-Coach-Review) for teaching caregivers evidence-based interaction strategies (responding, modeling, expanding, prompting) and input strategies (toy talk, recasting, READY talk; Cleave, Becker, Curran, van Horne, & Fey, 2015; Hadley et al., 2017b; Leech, Wei, Haring, & Rowe, 2018) to maximize the transition to syntax and decontextualized language and establish positive developmental trajectories.

There is consistent evidence that caregiver-implemented interventions change caregiver use of strategies for supporting language and improve children's language outcomes (Roberts & Kaiser, 2011; Barton & Fettig, 2013; Hampton & Kaiser, 2016; Heidlage et al., submitted; Law, Garrett, & Nye, 2004). Caregiver-implemented EMT has been used at high levels of fidelity in five RCTs conducted by our team including one study (Roberts & Kaiser, 2015) that specifically targeted children with RE delays at risk for DLD. We have developed and tested a specific model for coaching caregivers (Roberts, Kaiser, Wolfe, Bryant, & Spidalieri, 2014) (Teach-Model-Coach-Review) which will be used to teach caregivers EMT interaction strategies incorporating four primary linguistic input strategies: (a) expansions, (b) toy talk, (c) recasting, and (d) R.E.A.D.Y. talk. The effectiveness of expansions and recasts is well-established (Cleave et al., 2015; Levickis, Reilly, Girolametto, Ukoumunne, & Wake, 2014), and these are core features of responsive language input in EMT.

Toddlers with RE language delays are an understudied population with elevated risk for persistent DLD. These children are likely to need intervention for both vocabulary and syntax, but intervention on early syntax has been overlooked. This gap must be filled because grammatical abilities are needed to communicate clearly about events that happen outside of the immediate situation. Thus, there is a critical need to develop and test early, developmentally focused interventions to reduce the cascading negative effects of DLD. We propose to examine the effects of EMT-SF, a caregiver-implemented intensive treatment, provided before age 4, on growth in lexical and sentence diversity, language outcomes on standardized measures of vocabulary and grammar, and use of decontextualized language at age 4. The EMT-SF intervention is hypothesized to accelerate language growth by teaching caregivers how to engage in responsive interactions, to model noun subjects and semantically specific words in simple sentences, and to adapt these strategies as their children transition from learning and producing words to sentences to using their language skills to talk about decontextualized topics. The intervention is also expected to prevent the onset of additional risks associated with DLD. The study will provide a much-needed description of the transition from words to grammar that will inform scientific understanding of the progression of DLD (Dollaghan, 2013) and assess the potential benefit of intervening during this optimal developmental window for children with cumulative risk for DLD.

2.3 RISK/BENEFIT ASSESSMENT

2.3.1 KNOWN POTENTIAL RISKS

There are no obvious serious risks for any of the participants. The time involved in the study for families assigned to the EMT-SF intervention group is considerable; however, the perceived benefits outweigh the costs to families in personal time. Furthermore, children in both groups may continue to receive community-based intervention according to their Individualized Family Service Plan — the current standard of care or through private speech-language therapists. The EMT-SF intervention is an additional and not a replacement intervention; as such, children do not have to forgo other treatments to participate in the study. Thus, the only potential risk to study participants associated with this project is the risk of confidentiality being violated. There are safeguards in place, as described below, to protect against breach of confidentiality. Overall, the risks associated with this study are very low.

2.3.2 KNOWN POTENTIAL BENEFITS

Benefits to caregivers who participate in the intervention are expected to include: (a) improved strategies for promoting their child's language and communication skills and (b) more accurate perceptions of the child's language and communication abilities. In our previous research studies, caregivers have rated our training as helpful to them and their children; they have reported high levels of satisfaction with the training protocol. Caregivers in both groups will have the benefit of reviewing periodic language assessment results. After each major assessment period when standardized tests are administered (36, 42, and 49 months of age) all caregivers will receive a comprehensive report describing their child's language progress.

2.3.3 ASSESSMENT OF POTENTIAL RISKS AND BENEFITS

The proposed intervention is designed to accelerate language growth and sustain developmental momentum to close gaps between children at risk for DLD and their peers by targeting key transitions in language development and by expanding the role of caregiver education and coaching into the preschool years. This intervention has the potential to transform preschool service delivery models, to strengthen the spoken language foundation needed for school success, and prevent or reduce the long-term risks of DLD. Thus, the proposed research has the potential to improve the current standard of care for children with DLD and minimal risks to participants.

3 OBJECTIVES AND ENDPOINTS

The primary objective of the study is to determine if an intervention (EMT-SF) addressing both interactional processes and linguistic input, delivered by caregivers and interventionists over the 18-month period when critical vocabulary and early syntax are developing, can improve global language outcomes at 49 months in a population of children with clinically significant receptive-expressive language delays.

The 49-month endpoint was selected because this is the age at which language development trajectories are stabilizing and DLD can be established. Age 4 also represents the point at which a diverse vocabulary and basic syntax are well-established in typically developing children and support the transition to complex syntax, pragmatic language use, and learning in more decontextualized situations. For these reasons, language abilities at 49 months are strong predictors of later social and academic functioning. Endpoints for key objectives of the study are shown in Table 2 below.

A successful clinical outcome would be a difference in standard score on the Core Language Score (CLS) of the Clinical Evaluations of Language Fundamentals-Preschool, 2nd Edition (CELF-P2; Wiig, Secord, & Semel, 2004). The CLS is made up of three "core" subtests (i.e., Sentence Structure, Word Structure, Expressive Vocabulary) that together are the most discriminating and clinically sensitive in identifying DLD. CLS standard scores on the CELF-P2 range from 45 to 155, with scores below 70 indicating severe DLD (i.e., percentile ranks < 2), scores between 71 and 77 indicating moderate DLD (i.e., percentile ranks between 3 and 6), and scores between 78 and 81 indicating mild DLD (i.e., percentile ranks between 7 and 10 (pp. 78, 170-171). An increase on the CLS of 6 standard score points or greater would reflect a

change in growth rate above and beyond the average growth rate for same-age peers. An increase of 6 standard score points could also be sufficient for changing a child's classification from a moderate to mild severity rating (e.g., 72 to 78), or from mild severity to within the average range (e.g., 80 to 86).

Table 2. Objectives, Endpoints, and Justifications

Objective Number	Research Question	Outcome	Outcome Time Point	Justification
1.1	What are the long-term effects of the intervention on overall language skills at 49 months of age?	CELF-P2 Core Language Score at 49 months	49 months	The age at which persistent developmental language disorder (DLD) can be reliably identified.
1.2	What are the long-term effects of the intervention on growth in vocabulary skills during intervention (from 30 to 49 months)?	Slope of total number of different spoken word (NDW) roots from the Language Sample measured every 3 months from 30 to 49 months	Intercept at 49 months from a multi-level growth curve (30, 33, 36, 39, 42, 45, 49 months)	Collecting data every 3 months provides a more stable estimate of vocabulary growth during intervention; 49 months of age is the age at which developmental language disorder (DLD) can be reliably identified.
1.3	What are the long-term effects of the intervention on growth in grammar skills during intervention (from 30 to 49 months)?	Slope of a latent variable: • Number of different verbs • Number of different subjects • Tense/agreement productivity • Index of productive syntax • Clauses per utterance from the Language Sample measured every 3 months from 30 to 49 months	Intercept at 49 months from a multi-level growth curve (30, 33, 36, 39, 42, 45, 49 months)	Collecting data every 3 months provides a more stable estimate of grammar growth during intervention; 49 months of age is the age at which developmental language disorder (DLD) can be reliably identified.
2.1	What are the short-term effects of the Phase 1 vocabulary intervention on child vocabulary?	Latent Variable: • PPVT standard score • EVT standard score At 36 months	36 months	The age at which the first phase of intervention is completed.
2.2	What are the short-term effects of the Phase 2 grammar intervention on child grammar?	Latent Variable: • SPELT-P2 raw score • TEGI raw score At 42 months	42 months	The age at which the second phase of intervention is completed.
2.3	What are the effects of the intervention on child communication and interaction in real world settings at 49 months of age?	Focus on the Outcomes of Children Under Six (FOCUS) Raw Score	Intercept at 49 months	The age at which persistent developmental language disorder (DLD) has an increased impact on everyday interactions.
2.4	What are the effects of the intervention on rate of diagnosis of developmental language disorder?	Diagnosis of a developmental language disorder based on SPELT-P2 and TEGI raw scores	49 months	The age at which persistent developmental language disorder (DLD) can be reliably identified.

Objective Number	Research Question	Outcome	Outcome Time Point	Justification
2.5	What are the effects of the intervention on decontextualized language (DL)?	Total number of utterances that contain decontextualized language from a language sample at 49 months	49 months	The age at which decontextualized language can be accurately assessed.
2.6	What are the effects of the intervention on narrative skills?	Information score from the Renfrew Bus Story at 49 months	49 months	The age at which narrative skills can be accurately assessed.
2.7	What are the effects of the intervention on internalizing behaviors?	Child Behavior Checklist Internalizing T score at 49 months	49 months	The age at which internalizing behaviors are associated with language delays.
3.1	Do baseline receptive language skills (PLS-AC standard score) moderate intervention outcomes?	Same as the 1.1, 1.2, and 1.3	Same as the 1.1, 1.2, and 1.3	The age at which persistent developmental language disorder (DLD) can be reliably identified.
3.2	Does biological sex moderate intervention outcomes?	Same as the 1.1, 1.2, and 1.3	Same as the 1.1, 1.2, and 1.3	The age at which persistent developmental language disorder (DLD) can be reliably identified.
3.3	Does family history of language impairment moderate intervention outcomes?	Same as the 1.1, 1.2, and 1.3	Same as the 1.1, 1.2, and 1.3	The age at which persistent developmental language disorder (DLD) can be reliably identified.
3.4	Does presence of disruptive behaviors at baseline moderate intervention outcome?	Same as the 1.1, 1.2, and 1.3	Same as the 1.1, 1.2, and 1.3	The age at which persistent developmental language disorder (DLD) can be reliably identified.
3.5	Does caregiver use of child-directed, developmentally appropriate, target language mediate intervention outcomes?	CELF-P2 Core Language Score at 49 months	49 months	The age at which persistent developmental language disorder (DLD) can be reliably identified.
3.6	Does early child growth in vocabulary and grammar during intervention mediate outcomes? (see Figure 4)	CELF-P2 Core Language Score at 49 months	49 months	The age at which persistent developmental language disorder (DLD) can be reliably identified.

4 STUDY DESIGN

4.1 OVERALL DESIGN

A multi-site, phase 2, randomized clinical trial will be used to determine whether communication support strategies are effective for improving language outcomes in children with emergent developmental language disorder. The central hypothesis is that the Enhanced Milieu Teaching-Sentence Focus (EMT-SF) intervention will result in greater vocabulary and grammar outcomes for children in the treatment group.

At study entry, 108 children with emergent developmental language disorder (DLD) at 30 months of age will be randomly assigned 1:1 to either the EMT-SF treatment condition or a Business as Usual (BAU) control group. The control group is necessary to determine the efficacy of the EMT-SF intervention. The EMT-SF group is necessary to evaluate the effects of systematically teaching caregivers to use these strategies. Because all children in the study have language delays that will make them eligible to receive the early intervention services through the state early intervention program, children in both experimental conditions will receive state-provided community-based intervention according to their Individualized Family Service Plan — the current standard of care or from private speech language therapy providers. Children in the EMT-SF condition will receive an additional 18-19 months of virtual interventionist plus caregiver-implemented intervention sessions (depending on age of entry into the study). Children in both groups will be assessed at the start of the study and every 3 months until the child is 49 months old. Our goal is to enroll all children around 30 months of age and provide a minimum of 60 of the planned 66 sessions of intervention to each child in the treatment condition; however, variability in age at study entry (e.g., 30 months), intervention dosage, and number of assessment data points will be addressed in the statistical analysis.

4.2 SCIENTIFIC RATIONALE FOR STUDY DESIGN

A randomized clinical trial of the proposed intervention is needed to determine the effectiveness of the combined caregiver-plus-interventionist approach to teaching vocabulary and early syntax over an extended period in early childhood. The main goal of this study is to determine if such an intervention is sufficient to improve language outcomes, ameliorate significant early developmental language delays, and improve the trajectory of language development in children with early receptive and expressive delays indicative of emergent developmental language disorder.

Young children with significant receptive and expressive language delays, cognitive skills in the typical range, and no other identified disabilities will participate with their caregivers. Selection of child participants meeting these criteria ensures that the population meets the criteria for significant early receptive and expressive language delays, demonstrates cognition within the typical range, and exhibits no clinically significant symptoms of ASD or another disability that might account for their language delay. Children who meet these criteria will be selected following the screening procedures described below.

Once participants consent, caregiver-child dyads will be randomly assigned to experimental conditions (EMT-SF or Control). At study entry, 108 children with emergent developmental language disorder (DLD) at 30 months of age will be randomly assigned to either the Enhanced Milieu Teaching-Sentence Focus (EMT-SF) treatment condition or a control group.

A stratified randomized clinical trial design will be used with site (Vanderbilt, Northwestern) and biological sex (boys, girls) as stratification factors and EMT-SF or business-as-usual (BAU) control as a randomization factor. Given that language ability is likely to vary among participants, we use biological sex as a stratification factor because biological sex is categorical and related to the development of language abilities.

Children will be assessed every 3 months for the duration of the intervention (18-19 months; total of 7 assessment time points). The control group is necessary to determine the extent to

which the targeted communication support strategies are present in existing caregiver-child interactions. The EMT-SF group is necessary to evaluate the effects of systematically teaching caregivers to use these strategies. Further, we assume there will be some changes in child language as a result of developmental processes; thus, comparing children who receive the EMT-SF treatment to a BAU control group allows us to determine the extent to which change in the treatment group exceeds what would occur as a result of development over the period of the treatment supported by business-as-usual interventions in the community. In addition, the majority of measures in this study are norm referenced; thus, we are able to compare actual language at specific age points to the expected language abilities at those time points. By examining growth in many of the proposed analyses, we are able to determine if the rate of change in language development differs for the treatment and control groups.

4.3 JUSTIFICATION FOR DOSE

Caregiver/child dyads in the EMT-SF intervention group will receive a total of 66 virtual intervention sessions over an 18 to 19-month period beginning when the child is 30 months old and continuing until the child is 49 months old. The length of time, treatment targets, and intensity of treatment were determined based on results from our initial RCT with this population. Findings from this RCT (Roberts & Kaiser, 2015; Hampton et al., 2017; Hadley, et al., 2018) and a meta-analysis examining longitudinal language outcomes for young children with receptive and expressive delays (Fisher, 2017) suggested that more intensive intervention (greater dosage, including both vocabulary and early syntax targets, and treating over a longer period of development) to promote the growth in core vocabulary and syntactic skills is needed to adequately prepare young children with DLD for successful school entry. The third phase of the EMT-SF intervention was included in light of longitudinal findings suggesting that short-term, vocabulary-focused EMT intervention were not sufficient to significantly improve children's narrative skills compared to the control group (Cunningham, Quinn, Kaiser, & Roberts, 2016). Thus, explicit caregiver instruction is included in the third phase to teach caregivers new strategies for modeling language and asking inferential questions during book reading and supporting decontextualized talk in conversational discourse that is similar to the academic talk children experience in preschool and kindergarten classrooms. Thus, Phase 3 focuses on two new contexts (shared book reading, decontextualized conversations) and teaches caregivers additional strategies for language modeling and question asking. The choice of contexts and linguistic strategies are based on meta-analytic findings about the effectiveness of caregiver-implemented dialogic reading interventions in building vocabulary (Mol, Bus, de Jong, & Smeets, 2008; Dunst, Williams, Trivette, Simkus, & Hamby, 2012) and the empirical findings from Rowe and colleagues (Rowe, 2012; Leech et al., 2018) demonstrating the importance of decontextualized caregiver-child talk in supporting early academic skills.

In order to address the need for intensive intervention, we have structured the 18-month intervention corresponding to differential emphasis on specific caregiver-child interaction strategies and contexts, treatment targets, and language input strategies. We will tailor the delivery of the intervention in response to children's emergent language abilities throughout the intervention. (A complete description of the intervention and tailoring criteria is in section 6.1.3.)

4.4 END OF STUDY DEFINITION

A participant is considered to have completed the study when the child is approximately 49 months of age.

4.5 INCLUSION CRITERIA

In order to be eligible to participate in this study, an individual must meet all of the following criteria:

1. Child that is between 30 and 31 months of age (all races and biological sexes included).
2. Parent or caregiver (mother, father, grandparent) willing to participate in caregiver training over the full 18-19 months of the study
3. Normal non-verbal cognitive abilities
 - a. Visual Reception (nonverbal cognition) subscale t-score score of 35 or greater on *Mullen Scales of Early Learning (MSEL)* (within 1.5 SD of the mean for typically developing same age peers);
4. Receptive and expressive language delay:
 - a. *PLS-5* Auditory Comprehension (receptive language) standard score of 95 or less (1 SD or more below the mean for typically developing same-age peers) AND
 - b. *PLS-5* Expressive Communication (expressive language) standard score of 80 or less (1.3 SD or more below the mean for typically developing same-age peers), OR Expressive language scores on the MacArthur-Bates Communicative Development Inventory (CDI) below $\leq 10^{\text{th}}$ percentile (1.25 SD or more below the mean for typically developing same-age and same-sex peers) on 2 of the following 3 sections:
 - i. CDI Words produced
 - ii. CDI Mean 3 Longest
 - iii. CDI Complexity
5. English as the only language spoken to the child in the home
6. Provision of signed and dated informed consent form
7. Attempts to imitate 10 words (per parent report), given imitation skills are a prerequisite skill of the intervention.

4.6 EXCLUSION CRITERIA

Children who meet any of the following criteria will be excluded from participation in this study:

1. Diagnosis of a neurodevelopmental disability other than DLD (e.g., Down syndrome, ASD, intellectual disability).
2. Caregiver report of a major medical condition (e.g., seizures, cancer, stroke, traumatic head injury, cleft lip/palate, cerebral palsy, legally blind, any genetic diagnosis associated with neurodevelopmental disability).
3. Caregiver report of hearing impairment or audiological testing indicating hearing thresholds $\geq 20\text{dB}$.
4. Caregiver report or direct observation of any problems chewing, sucking through a straw, or blowing bubbles.

4.7 LIFESTYLE CONSIDERATIONS

N/A

4.8 SCREEN FAILURES

Screen failures are defined as participants who consent to participate in the clinical trial but are not subsequently randomly assigned to the study intervention or entered in the study. A minimal set of screen failure information is required to ensure transparent reporting of screen failure participants, to meet the Consolidated Standards of Reporting Trials (CONSORT) publishing requirements and to respond to queries from regulatory authorities. Minimal information includes demography, screen failure details, eligibility criteria, and any serious adverse event (SAE).

Participants will be screened in two stages. First, participants will be screened over the phone or online survey. A designated staff member will screen all participants by asking caregivers a standard set of eligibility questions, including: (1) age of the child; (2) if caregivers have language concerns about the child or if the child is receiving speech-language therapy; (3) child's hearing ability; (4) additional diagnoses or medical concerns; (5) if a parent/caregiver is willing to participate in intervention; and (7) if the child only hears English spoken in the home. Second, if the child and family meet initial criteria during the phone screen, the family will complete the MCDI-WS survey. If the child meets eligibility criteria in the domains of words produced, mean 3 longest and complexity, then a virtual screening visit will be scheduled to confirm eligibility (Preschool Language Scale-5th Edition, PLS-5; Telemedicine-based ASD Evaluation Tool for Toddlers and Young Children, TELE-ASD-PEDS; ADI-R, Mullen Scales of Early Learning, MSEL – Visual Reception). The PLS-5 will be used to measure children's overall expressive and receptive language abilities. The MSEL-visual reception will be used to assess nonverbal cognitive abilities. The TELE-ASD-PEDS will be used to confirm the absence of ASD. The TELE-ASD-PEDS is a telehealth screening tool designed to identify children under 36 months of age who may be on the autism spectrum (ASD). If there are still clinical concerns about autism after the child has passed the TELE-ASD-PEDS, the Autism Diagnostic Interview (ADI-R) will be conducted with the parent to confirm the absence of autism. All assessments will be conducted by the caregiver with instructions and coaching from an assessor trained to research fidelity on each assessment prior to the beginning of the study.

Individuals who do not meet the criteria for participation in this trial (screen failure) because of testing scores will not be rescreened.

4.9 STRATEGIES FOR RECRUITMENT AND RETENTION

The sample will consist of 108 children with receptive and expressive language delays and their families meeting the aforementioned criteria. Families will be randomly assigned to a treatment or control condition. Northwestern University and Vanderbilt University will be the two sites responsible for recruitment. Children will be recruited primarily through Early Intervention providers in the United States, pediatrician offices, advertisements, and early childhood education providers. Online advertising will also be utilized if necessary. Intervention providers will distribute study flyers to the parents and/or guardians of children with receptive/expressive language delays at 30 months of age. We will target younger children before they reach entry

age for the study, to ensure that we complete screening and pretesting as close to 30 months of age as possible. Interested families will contact their corresponding site (Vanderbilt or Northwestern) to find out more information and complete a phone eligibility screening.

Monthly recruitment benchmarks will be established to continuously monitor recruitment and retention. If recruitment falls below the benchmark for more than 2 consecutive months, additional recruitment strategies (e.g., recruitment through advertising such as magazines, Facebook) will be employed. To meet the recruitment goals of the study (108 children; 54 per clinical treatment site), we will recruit a minimum of 2 participants per month for 27 months. We have successfully recruited this population at the Vanderbilt site for a previous RCT (92 caregiver-child dyads in 3.5 years) and 120 young children with ASD and their caregivers across the Vanderbilt and Northwestern sites in a current RCT. We have established connections for recruitment and have processes in place for efficient screening and study entry.

Persons of both biological sexes and all races and ethnic backgrounds may be included in this study. No participants will be excluded on the basis of biological sex or ethnic background; however, the capacity to deliver the proposed intervention in languages other than English does not exist at this time.

It is anticipated that most of the adult participants ($n = 108$) involved in the study will be mothers or female caregivers (aunts, grandmothers), although a small number of fathers (less than 10%) have participated in previous studies. In past research, approximately 5% of adult participants have been non-parental primary caregivers (e.g., grandparent, aunt). Only one caregiver from each family will participate in the training sessions; however, adult language measures will be collected from both biological parents, when possible. Caregivers in previous studies have come from lower- to upper-middle socioeconomic backgrounds and the same population is expected in this study. Typically, populations of caregivers in previous studies have reflected the racial composition of the greater metropolitan recruitment areas.

Vulnerable populations included in this research are children with receptive and expressive (RE) language delays at 30 months of age at baseline. This population is specifically included in order to determine the best early intervention practices for children with RE language delays and their caregivers. Children with severe RE language delays are less likely to spontaneously recover compared to children with only expressive language delays, and thus, children with RE delays are likely to experience persistent developmental deficits in the language domain. Children at 30 months of age were selected because cognitive skills and language skills can be reliably assessed at this age. In addition, this is a critical point in early language development when developmental gaps between children with RE language delays and their typically developing peers are likely to increase. In the absence of intervention, slow vocabulary growth is likely to impede grammatical development. Thus, early RE language delays are predictive of future language abilities.

Maintaining participation in the study. We will use several strategies that have been successful in maintaining families in studies and achieving a very low rate of attrition in past studies. All fully enrolled families (who pass screening) will receive an iPad, case, tripod, and wireless headset to help with data collection, which include video recorded sessions in their home. Families will also complete an internet speed test and if their internet speeds are not at the minimum required to support a video call, a cellular-enabled iPad will be given to the family for the duration of the study. If the family completes the final visit at 49 months, they will be

allowed to keep the iPad and all accessories. We will provide written reports of children's progress after the child's 30, 36, 42, and 49 month testing visits. Participating caregivers will also be offered clips from their study sessions on a password protected USB drive. In addition, all participants will receive equivalent sets of toys and books at 30, 33, 36, 39, 42, and 45 months of age (approximately \$165 at 30 months, \$80-110 at 33, 36, 39, 42, and 45 months). All participants will receive \$25-100 for completing each study assessment timepoint (\$100 at 30 months, \$2525 at 33 months, \$50 at 36 months, \$25 at 39 months, \$50 at 42 months, \$25 at 45 months, and \$100 at 49 months; amounts are based on the time required to complete the assessments). In addition, participants will also receive an additional \$25 for completion of online surveys at 30, 33, 36, 39, 42, and 49 months. Given that attrition is more likely in the control group and that the intervention group is receiving \$6,600 (66 sessions X \$100 dollars) worth of free intervention services, the control group will receive additional incentives (\$200 for completing all study activities at 36 months, \$200 for completing all study activities at 42 months, and \$200 for completing all study activities at 49 months). This corresponds to a total payment (excluding toys) of \$550 for participants in the treatment group and \$1,150 for participants in the control group over the 18-19 months of the study.

Control Payment Structure:

	30 mon.	33 mon.	36 mon.	39 mon.	42 mon.	45 mon.	49 V1	49 V2	Total
Assessments	\$100	\$25	\$50	\$25	\$50	\$25	\$50	\$50	\$375
Surveys	\$25	\$25	\$25	\$25	\$25	\$25	\$25		\$175
Data bonus			\$200		\$200			\$200	\$600
Total	\$125	\$50	\$275	\$50	\$275	\$50	\$75	\$250	\$1,150

Treatment Payment Structure:

	30 mon	33 mon	36 mon	39 mon	42 mon	45 mon	49 V1	49 V2	Total
Assessments	\$100	\$25	\$50	\$25	\$50	\$25	\$50	\$50	\$375
Surveys	\$25	\$25	\$25	\$25	\$25	\$25	\$25		\$175
Total	\$125	\$50	\$75	\$50	\$75	\$50	\$75	\$50	\$550

5 STUDY INTERVENTION

5.1 STUDY INTERVENTION(S) ADMINISTRATION

5.1.1 STUDY INTERVENTION DESCRIPTION

All families participating in the study assigned to either the treatment or control group will likely participate in Community-Based Intervention Services (CIS)- educational, speech-language, and other interventions provided in the community. If a child is not receiving CIS as part of the state early intervention program, we will refer them to the statewide early intervention program. Children in both groups are eligible to receive community-based intervention according to their Individualized Family Service Plan or Individualized Education Plan— the current standard of care. Children must have a 30% delay or greater in IL or a 40% delay or greater in TN in

communication skills as measured by a state-approved eligibility assessment tool. The EMT-SF intervention is an additional intervention and not a replacement for CIS; as such, children do not have to forgo other community-based treatments to participate in the study. We will ask caregivers to report their child's community interventions at each 3-month assessment (6 times across the 18-19 months of the study); we will analyze the summarized data to determine if there are significant differences in use of community services between groups.

Caregivers in the **treatment group** will participate in CIS, receive toys, books, and printed intervention materials and receive 66 EMT-SF virtual intervention sessions over 18 to 19 months (to account for cancellations). They will be taught the intervention strategies individually using a manualized protocol (Teach-Model-Coach-Review). Table 6 (see section 6.1.2) provides a summary of the intervention phases, caregiver strategies, child treatment targets, and criteria for determining advancement through the intervention phases. Caregivers in the treatment group will receive toys and all printed intervention instructions and payment for assessments on the same schedule as caregivers in the control group. Printed intervention instructions will include a description of each intervention strategy with specific ideas about how to use the target intervention strategy with toys and books.

Caregivers in the **control group** will participate in CIS, receive the same printed intervention instructions, toys, and books at the same intervals as the treatment group but will not receive EMT-SF intervention.

The only differences between groups is: (a) the virtual EMT-SF intervention sessions and (b) the incentive structure as outlined in section 4.9. The control group participants will receive all printed intervention instructions and toys at the same intervals as the treatment group. Both groups will receive equivalent sets of toys and books, which includes \$165 worth of books and toys at 30 months and \$85-\$110 at 33, 36, 39, 42, and 45 months of age during their assessment visits in the clinic. See Table 3 for a summary of study components by experimental condition.

Table 3. Summary of Study Components by Experimental Condition

	Treatment Group	Control Group
Intervention Materials	Books and toys every 3 months	Books and toys every 3 months
Intervention Instructions	Printed handouts	Printed handouts
Coaching on Intervention Strategies	66 virtual coaching sessions	0 coaching sessions
Community-Based EI Services	As specified according to the child's IFSP	As specified according to the child's IFSP
Total Amount of Incentives	\$550 plus iPad	\$1,150

5.1.2 DOSING AND ADMINISTRATION

The EMT-SF intervention will be delivered to individual caregiver-child dyads by an experienced speech-language pathologist or early childhood special educator who will serve as the interventionist. A total of 66 virtual sessions will be delivered over 18-19 months, depending on cancellations. The intervention will consist of three phases and will follow the schedule shown in Table 4 Overview of the EMT-SF Intervention and Child Treatment Targets.

Intervention Phase Overview. Each phase of the intervention corresponds to the child's individualized targets from the sentence-focused sequence (core vocabulary, sentences, and decontextualized language) with caregiver language strategies designed to support those targets. The sentence-focused sequence is a lexically based sequence of treatment targets designed to address the common verb vocabulary and grammatical deficits observed in DLD (Conti-Ramsden & Jones, 1997; Fletcher & Peters, 1984; Loeb, Pye, Redmond, & Richardson, 1996; Oetting, Rice, & Swank, 1995; Olswang, Long, & Fletcher, 1997; Rice, Buhr, & Nemeth, 1990; Watkins, Rice, & Moltz, 1993; Windfuhr, Faragher, & Conti-Ramsden, 2002). The sentence-focused targets will be individualized based on each child's language level at baseline and modified in response to their progress during the intervention.

Phase 1 provides intensive caregiver training (two sessions per week) in evidence-based interaction and language modeling strategies (responding, modeling, expanding, and prompting) to support children's core vocabulary development. Phase 1 draws upon the effective Teach-Model-Coach-Review (TMCR) process. Thus, the primary objective of Phase 1 is caregiver mastery of foundational EMT-SF interaction and language modeling strategies. Advancement to Phase 2 is based on caregiver mastery of foundational strategies as well as child achievement of core vocabulary and production of description and location sentences. In our previous caregiver-implemented, vocabulary-focused RCT for children with RE delays, caregivers were able to master the foundational EMT strategies in 24 sessions (Roberts & Kaiser, 2015; range 16-24 sessions). Caregiver training will switch to one session per week after the initial three months or 24 sessions. It is expected that the one time per week stage may span the second portion of Phase 1 as well as Phase 2 for some children.

In Phase 2 (action sentences), caregiver training will focus on tailoring the foundational strategies (e.g., model, expand, prompt) to the child's individualized targets from the sentence-focused sequence: action sentences. The duration of Phase 1 and 2 is expected to vary from 10 to 18 months, it typically takes younger, language-matched children 9 to 15 months to progress from early word combinations to diverse sentences (Hadley et al., 2018).

Phase 3 will focus on teaching caregivers to adapt the strategies to their child's changing developmental needs and to support verbal interaction in more challenging decontextualized conversations. All caregivers will receive Phase 3 decontextualized instruction during the final 6 months of the study. Caregivers will learn how to support longer, multi-turn conversations, to create opportunities for decontextualized conversations about past or future events using evidence-based R.E.A.D.Y talk strategies (Leech et al., 2018), to model narratives and use open-ended questions in the context of shared book reading, play and snack. Sessions in Phase 3 are less frequent because the EMT-SF strategies are being adapted and maintained, rather than established. Prior research has demonstrated that caregivers can learn the new Phase 3 strategies with brief instruction.

Specific criteria for caregiver and child advancement through the phases and the sequence of treatment targets are provided in Table 4 and in the section Delivery of the EMT-SF Intervention. Examples of possible variations in individual children's treatment targets within and across intervention phases as a function of caregiver mastery of EMT strategies and children's rate of treatment progress are provided in Table 5.

Child Target Selection. The intervention begins by choosing specific vocabulary and sentence targets based on the child's spontaneous communication skills obtained from the *PLS-5*,

caregiver-child interaction, and language sample at entry to the study. Targets will be modified based on child progress throughout the intervention (see Table 4 and description below). Early vocabulary targets include spontaneous use of diverse nouns as well as core prepositions, adjectives, and verbs (Phase 1 targets; core vocabulary), followed by intransitive verbs and location and description sentences that encode simple states. Action sentences with third person subjects and process verbs are then targeted to promote children's ability to encode more complex events in linguistically specific ways (Phase 2 targets; action sentences). Finally, decontextualized language will be targeted to support verbal interaction in more challenging decontextualized conversations. The target sequence is guided by linguistic theories of event structure in simple sentences and empirical evidence supporting this developmental sequence (Bowers, 1993; Pustejovsky, 1991, Ramchand, 2008; Hadley, McKenna, & Rispoli, 2018). In Phase 2, caregiver training will focus on tailoring foundational strategies to support the child's sentence targets. If children do not progress through the target sequence by 42 months, their targets will continue in Phase 3.

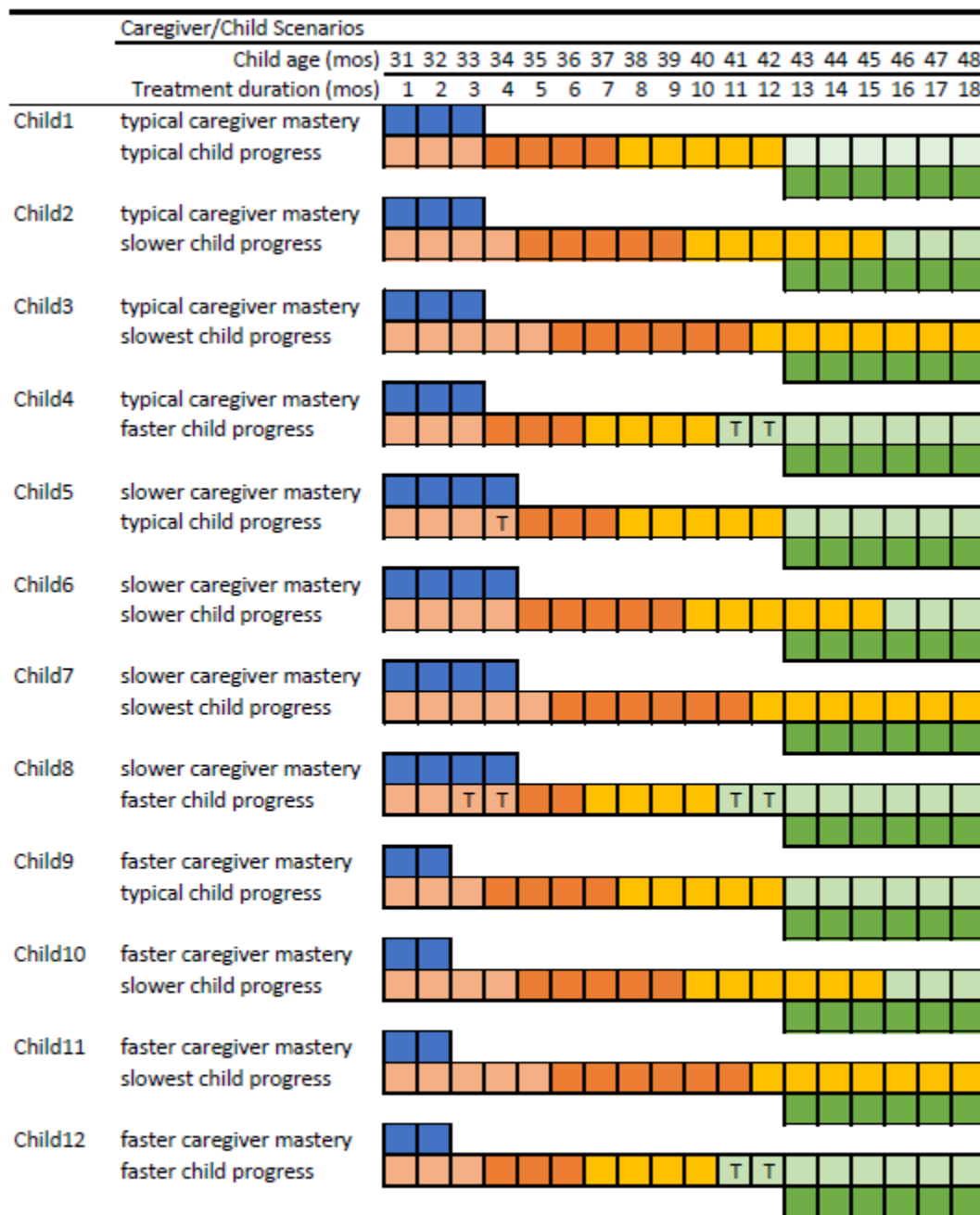
Table 4: Overview of the EMT-SF Intervention Phases, Caregiver Strategies, Child Targets, and Criteria to Advance

	Phase 1 Core Vocabulary Location/Description Sentences	Phase 2 Action Sentences	Phase 3 Decontextualized Language
Contexts	Routines, play, books	Routines, play, books	Play, books, conversations
Frequency	2 times per week Sessions 1-24 1 time per week Sessions 25-48	1time per week Sessions 49-60	1 time per month Sessions 61-66
Duration	6 months (planned) 5-9 months (min, max)	6 months (planned) 5-9 months (min, max)	6 months (planned) 6 months (min, max)
Caregiver Emphasis	Foundational EMT-SF Strategies	Maintenance of foundational strategies, tailoring strategies (modeling, expansion, prompting) to Child Targets	Adapt strategies for Decontextualized topics
Child Emphasis	Spontaneous Communication Rate and Child Sentence-Focused Targets	Child Sentence-Focused Targets	
Targets	Caregiver Strategies: Responsivity Pacing Model Expand Prompt Child Vocabulary Targets: Object names / Nouns (N) Location words / Preposition (P) Description words / Adjective (A) Action words / Verb (V) Child Sentence Targets: Location / Entity-location (N-P) Description / Entity-state (N-A)	Caregiver Strategies Model Expand Prompt Child Vocabulary Targets: Verbs: Intransitive (general) Intransitive (specific) Transitive (specific) Child Sentence Targets: 3P Subjects (S), increase diversity Action / SV and SVO SV: Intransitive Process SVO: Transitive Process-(Result)	Caregiver Strategies cont. Child Targets: Sophisticated Vocabulary and Decontextualized Talk OR Phase 2 Child Vocabulary and Sentence Targets (con't)
Interaction Nonverbal	Follow child's interest Mirror/expand play Specific combinations Child as agent Single-step sequences	Doll as agent Multi-step sequences	
Interaction Responsivity	Respond to communication Paced turns	(cont.)	Build multi turn conversations
Language Modeling (Input	Model content-rich words (N) and core vocabulary targets (prepositions, adjectives, verbs)	Model Target Verbs and Action Sentence Types (SV, SVO)	Model R.E.A.D.Y talk: <i>Recall past</i> <i>Discuss future</i>

Quality)	Model Target Verbs and Location/ Description Sentence Targets		
	Expand Add 1-2 content words Talk about the toys Add a location (target) Add a description (target)	Expand (with toy talk) Add an action word (target) Recast (with toy talk) Give object its name	Expand Add explanations for words and ideas Broad recasts
Child Productions	Time Delay	Same strategy used with different child targets	Ask open ended questions to infer, predict, explain
	Prompt sequence	Same strategy used with different child targets; level of support changes over time	
Criteria to Advance	Caregiver mastery of foundational EMT-SF strategies: measured by a score of "Pass" for each strategy item(s) on a global rating scale. Measured during 2, consecutive 6-minute unsupported caregiver-child interactions during intervention sessions by the therapist. AND Child production of (a) ≥ 30 different words, and (b) ≥ 4 different core location words, and (c) ≥ 2 different core description words in order to move the child to a location and description sentence targets Measured during 3, consecutive 6-minute unsupported caregiver-child interactions during intervention sessions or during the most recent standardized language sample assessment.. AND Child production of (a) ≥ 5 different intransitive verbs, and (b) ≥ 10 different location/description sentences in order to move the child from a to action sentence targets. Measured during 3, consecutive 6-minute unsupported parent-child interactions during intervention sessions or the most recent standardized language sample assessment.	Caregiver mastery of EMT-SF strategies with updated language: measured by a score of "Pass" for each strategy item(s) on a global rating scale. Measured during 2, consecutive 6-minute unsupported caregiver-child interactions during intervention sessions by the therapist. AND Child production of (a) ≥ 10 different action sentences with 3P subjects, and (b) ≥ 2 different noun subjects Measured during 3, consecutive 6-minute unsupported caregiver-child interactions during intervention sessions or during the most recent standardized language sample assessment.	All caregivers will receive instruction on decontextualized language regardless of child progress on Phase 3 targets.

More information about changing child targets based on specific criteria is included in the "Monitoring Criteria to Advance" section.

Table 5. Examples of Possible Child Treatment Sequences



Treatment duration (in months) is indicated by colored bars. Advancement from early Phase 1 (EMT strategies, blue/child core vocabulary, light orange) to late Phase 1 (child location/description sentences, orange) is based on caregiver mastery of EMT strategies and child production of diverse core vocabulary. Advancement from late Phase 1 to Phase 2 (Child action sentences, yellow) is based on child production of diverse location/description sentences. If children produce diverse action sentences prior to 42 months, sophisticated vocabulary targets (light green) will be introduced prior to direct caregiver instruction on decontextualized language in Phase 4. T indicates therapist modeling of new targets prior to caregiver instruction.

Interventionists will be trained to criterion before the study begins, assigned to individual families at two clinical sites, and follow manualized procedures for teaching caregivers during each hour-long session using the TMC protocol shown in Table 6 Intervention Session Format below. Additional details for each phase of the intervention are presented in section 6.1.3 immediately below and fidelity monitoring procedures for delivery of the caregiver training are described in section 10.1.7.

Table 6. General Intervention Session Format

Session Elements	Definition
Probe (6-minutes)	6-minute caregiver-child interaction at the beginning of every session to see how the caregiver is using the strategies.
Teach (10 minutes)	The interventionist will review how the caregiver used the strategies during the probe including strengths and areas for improvement. The interventionist will then describe the target intervention strategy of the day including the rationale. She will give an example of the strategy in relation to the selected toys for the session and ask the caregiver to come up with another example to check understanding. When the caregiver is first learning a strategy, the interventionist will also play a standardized clip of an interventionist modeling the strategy with a child to demonstrate what it looks like. The interventionists will also help the caregiver plan for the session including which toys/objects to use and which activities to complete.
Model & Coach (30 minutes)	The caregiver practices the procedures and the interventionist provides the caregiver with constructive feedback or specific praise. The interventionist will also model at least 10 examples of target language for the caregiver. This may be during a time the child is solely interacting with the therapist over the screen or interspersed throughout the caregiver's practice with the child.
Review (10 minutes)	The interventionist summarizes how the caregiver used the target strategies, relates caregiver use of strategies to child communication during the session, and discusses how the strategies will be used at home.

Delivery of the EMT-SF intervention. A research interventionist will teach the caregiver and interact with the caregiver-child dyad during individual virtual sessions lasting approximately 60 min. Prior to the start of every session, the therapist will record a 6-minute caregiver-child interaction sample without any support to the caregiver. This 6-minute caregiver-child interaction session allows the therapist to see how the caregiver implements the strategies without any support. As such, this session will be used to determine child and caregiver progress and when the dyad is ready to change intervention phases. After this 6-minute caregiver-child interaction sample, the intervention sessions will include: (a) the interventionist describing the target intervention strategy (Teach), (b) the interventionist modeling language targets for the caregiver while interacting with the caregiver-child dyad with similar toys via telecommunications (Model),

(c) the caregiver practicing the strategy with the child with coaching from the interventionist (Coach), and (d) the interventionist reviewing the session with the caregiver, summarizing the session, answering questions and planning for independent implementation (Review) (see Table 6 above). Sessions will be conducted virtually during play, routines, and book reading interactions. The caregiver may use the toys and materials given as incentives for participation to both treatment and control families throughout the intervention period.

Phase 1 (Core Vocabulary & Location/Description Sentences). Caregivers will be introduced to the intervention strategies following the instructional sequence used in the previous RCT (see Table 6 above). In Phase 1, treatment sessions will initially occur two times per week (sessions 1-24) and then one time per week (sessions 25-48). The primary goals of Phase 1 is to teach caregivers the foundational EMT-SF strategies and increase children's rate of verbal communication with diverse words, and begin to combine nouns with core location and description words to form location and description sentences. Caregivers will learn nonverbal interactive strategies and verbal responsive interaction strategies to increase children's spontaneous communication and language learning opportunities. Responsive interaction provides the foundation for communicative exchanges across the three phases of the intervention, whereas language modeling strategies and language targets vary across the intervention phases. In the first phase, caregivers will learn to model and combine content or 'teaching words' in simple location and description sentences. Caregivers will also learn to prompt child productions when their child is requesting an object with expectant looks, choice questions, and imitation prompts. Once all strategies have been introduced, treatment sessions will focus on caregiver integration of all strategies. As children progress through Phase 1 targets, caregiver training will focus on tailoring EMT-SF strategies to model the verb vocabulary, and to expand and recast child utterances to highlight SF sentence targets with toy talk sentences (Cleave et al., 2015; Hadley et al., 2017a; Hadley et al., 2017b). Toy talk modifies adult language input in ways that increase subject diversity in adult input, and in turn, this input modification has been shown to accelerate children's grammatical growth.

The *core vocabulary* targets will include 6 early location words (prepositions; *in, out, on, off, up, down*) and 10 early descriptive words (adjectives; *open, closed, wet, dry, clean, dirty, hot, cold, empty, full*) incorporated into play and daily routines. To advance from these vocabulary targets to location and description sentence targets, the child must produce (a) at least 30 different words spontaneously; (b) at least 4 different core location words (prepositions), and (c) at least 2 different core description words (adjectives) during 18 minutes of caregiver-child interaction (6 minute interaction samples from 3 consecutive unsupported interactions) or during the most recent Language Sample (an observational assessment used at all assessment visits).

After 24 bi-weekly sessions (~ 3 to 4 months), sessions will occur weekly. In late Phase 1, some children may be ready to advance to core verb targets and to location and description sentence targets whereas others may not advance until after the first six months (see Table 5). Core verb targets in late Phase 1 will be 10 early action words (activity verbs; *eat, drink, play, sleep, cry, walk, run, dance, ride, do*) with simple event complexity (i.e., ACT) and other early developing intransitive verbs (e.g., *go, come, fall, get, break, cook, fit, jump, roll, spill, sit, stop*). Caregivers will receive a list of common intransitive verbs that can be incorporated into play and daily routines. In this phase, children are likely to produce action words in word combinations without a subject.

The rationale for location and description sentence targets in late Phase 1 is to strengthen children's mental representation of the subject position in sentences that do not require a verb. The description and location sentence targets combine the core vocabulary from early Phase 1 into two specific semantic relations: entity-location (e.g., *it/hat on*) and entity-state (e.g., *it/diaper wet*). These semantic relations share the following syntactic structure: a single argument in the subject position of a simple state predicate. Caregiver training will focus on modeling and expanding the target verbs and sentence types in short, grammatical input sentences. Coaching in late Phase 1 will focus on expanding children's production of nouns with core vocabulary targets (i.e., location or description words) (e.g., *cow* → *the cow is in*), adding nouns to children's production of core vocabulary (e.g., *fall* → *the cup fell*), and recasting third person pronoun subjects with noun subjects (e.g., *C he hungry* → *A The dog is hungry*). Through use of these strategies, subject diversity in caregiver input will be increased. To advance to Phase 2, the child must produce (a) at least 5 different intransitive verbs spontaneously; and (b) at least 10 different Phase 1 target sentences during 18 minutes of caregiver-child interaction (6 minutes interaction samples from 3 consecutive sessions) or during the most recent Language Sample. In addition, the caregiver must demonstrate mastery of the Phase 1 strategies. This will be measured by a score of "Pass/Fail" based on clinical judgment across 2 consecutive unsupported caregiver-child interactions. Meeting these criteria ensures that caregivers are fluent in the EMT-SF strategies and that children have the basic sentence structure needed for action sentences with later developing noun subjects and verbs encoding more complex event structure.

Phase 2 (Action Sentences). In Phase 2, verb targets are more lexically specific and will increase in event structure complexity (i.e., GO, BECOME, CAUSE) as children progress through the sequence of verb and sentence targets. Children's target verbs and sentences will be highlighted with natural prosody in short, grammatical phrases and sentences across all levels. Phase 2 focuses on expanding the transitive verb vocabulary and combining verb in action sentences with third person subjects. Caregivers will receive a list of common transitive verbs that can be incorporated into play and daily routines. Verb target examples include *put, make, give, take, feed, fix, push, pull, bounce, drop, knock, kick, pour, fill, dump*. Sentence targets in this phase increase in complexity and have two shared characteristics: a third person subject and a process verb (i.e., ACT, GO, BECOME, CAUSE). The target sentence types include three SV sentence types with intransitive process verbs: (a) entity-GO-location (e.g., *hat go on*), (b) entity-BECOME-state (e.g., *diaper got wet*), (c) Entity-ACTIVITY (e.g., *tower fall*) and (d) SVO sentences with transitive process verbs with or without a result phrase (e.g., *baby drinking juice; horse knock fence down*). Caregiver training will continue to focus on tailoring EMT-SF strategies for the child's updated target. Caregivers will model child verb and sentence targets in short, grammatical input sentences. Child sentences are likely to be in simplified, telegraphic form, providing opportunities for caregivers to recast child sentences into well-formed adult versions in the subsequent, responsive turn. To advance from Phase 2 to Decontextualized (Phase 3), the child must produce (a) at least 10 different action sentences spontaneously; and (b) at least 2 different nouns as subjects in sentences during 18 minutes of caregiver-child interaction (6 minutes from interaction samples from 3 consecutive sessions) or during the most recent Language Sample. Mastery of these targets is important insofar as the linguistic ability to produce diverse sentences, especially with noun subjects, will help children communicate about objects and events more effectively in decontextualized contexts. In addition, the caregiver must demonstrate mastery of the Phase 2 strategies. This will be measured by a score of "Pass/Fail" based on clinical judgment across 2 consecutive unsupported parent-child interactions.

Phase 3 – Decontextualized. The third and final phase of the intervention is the EMT-SF Decontextualized Phase. Sessions in the third phase will ideally take place on a monthly basis (if no sessions have been missed, requiring additional makeup sessions). All caregivers will be provided with ongoing instruction and support in adapting the interaction and language modeling strategies to the changing developmental needs of their children. Caregiver instruction in the final phase will focus on how to arrange the environment; to model, expand, and prompt children for more sophisticated language; and respond with the support needed to build multiple turn conversations in activities that often require comprehension and production of decontextualized talk. Thus, intervention activities in the EMT-SF Decontextualized phase will emphasize caregiver language input during shared book reading, play, and snack. If children have not progressed through all EMT-SF target levels, SF child targets will continue to be the focus. Caregiver modeling and prompts for child production of decontextualized talk will be tailored to the child's language level and prompts may be limited (i.e., 1 prompt in 15-min coaching segment). If a child has mastered all SF targets, caregivers will be taught to model more sophisticated vocabulary to promote vocabulary growth, to respond with broad recasts to promote grammatical growth, to model longer narratives and to prompt for child production of decontextualized talk more frequently (i.e., 3 or more prompts in 15-min coaching segment).

Caregivers will be introduced to new evidence-based strategies to model decontextualized language during caregiver-child interactions, a predictor of children's future vocabulary development and school success (Demir, Rowe, Heller, Goldin-Meadow, & Levine, 2015; Rowe, 2012; Snow, Burns, & Griffin, 1998). These language modeling strategies will be referred to as R.E.A.D.Y. talk (Leech et al., 2018). The first four letters of R.E.A.D.Y. stand for Recall the past, Explain new words and ideas, Ask questions, and Discuss the future. The letter Y stands for You can make a difference in your child's school readiness. Thus, R.E.A.D.Y. talk is intended to aid caregivers' memory for and motivation to use these language modeling strategies. These strategies first will be taught in shared book reading. Using the TMCR process, caregivers will learn to relate the story to past experiences (e.g., *Do you remember when we ...*), to explain word meanings from the story and to make inferences. Caregivers will also learn to use open-ended questions to prompt children for more complex language, to make inferences, to explain characters' feelings (e.g., *Why was the red hen mad?*) and to make predictions (e.g., *What do you think will happen next?*). In addition to practicing in shared book reading, caregivers will identify one other everyday activity (e.g., family meal, car ride) in which to practice modeling decontextualized talk about one topic at least one time each day outside the treatment sessions.

Monitoring Child Criteria to Advance. Child advancement criteria will be monitored by transcribing and coding child utterances. These samples will be taken from three unsupported 6-minute caregiver-child interaction sessions conducted at the beginning of the intervention sessions as well as transcripts from the Language Sample at the 33-month, 36-month, 39-month, 42-month, and 45-month assessment visits. All child utterances will be transcribed and then coded for frequency of child vocabulary and sentence targets. When the child meets the criteria, these data will be discussed on a weekly cross-site clinical meeting before the therapist moves to the next intervention phase.

Child progress during intervention will also be monitored by transcribing and coding child utterances. These samples will be taken from 6-minute unsupported caregiver-child interactions during the beginning of intervention sessions and 6-minute supported caregiver-child interactions during the coaching portion of intervention sessions. Interactions will be coded twice per month for month 1, 2, and 3, and once per month during the subsequent months for a total

of 21 observations. Staff members at Vanderbilt who are not naïve to experimental condition will transcribe and code the videos using Interact and then send blinded transcripts to coders at UIUC who are naïve to experimental condition. The UIUC coders will code the videos for frequency and diversity of child linguistic targets using SALT.

Monitoring Caregiver Implementation of the EMT-SF intervention. We expect caregivers will acquire and implement the EMT-SF procedures taught during the intervention given well-trained interventionists following manualized training procedures and weekly clinical supervision of the interventionists. In previous EMT studies with this population, caregivers achieved mastery of the Phase 1 EMT strategies within 24-30 sessions (Roberts & Kaiser, 2015; Roberts et al., 2014.) At the end of each intervention session, the interventionist will discuss how to apply the newly learned and previously learned intervention strategies on their own (Review component of training protocol). Once target words and sentence types are identified, caregivers will work with the interventionist to develop plans for modeling target words and sentences in the upcoming weeks.

Caregiver implementation will be monitored by scoring 6-minute caregiver-child interaction sessions conducted prior to the intervention sessions by coding the videos for intervention strategy use, reviewed monthly by the interventionist. In addition, caregiver implementation of the intervention strategies will be assessed at each major time point during a virtual caregiver-child interaction recorded via telecommunications.

Training and Monitoring Interventionists' Instruction of Caregivers. Interventionists who deliver the virtual intervention will be trained to criterion before the intervention begins by a master interventionist (Suzanne Thompson or Jen Nietfeld, Vanderbilt). Interventionists at both sites will receive initial training during a 2-day workshop at Vanderbilt prior to the start of the trial. The procedures to be used in training caregivers will be described in a comprehensive treatment manual. Each interventionist will complete practice sessions with children and caregivers and receive coaching and feedback from Mrs. Thompson, Mrs. Nietfeld., and Dr. Hadley. Criterion for fidelity for interventionists will include demonstrating research fidelity on the primary child intervention EMT-SF in three practice sessions with a child similar to the target population of children enrolled in the study, and demonstrating fidelity in delivering the individual components of the caregiver training provided in intervention sessions. The fidelity instruments used throughout the study will be used in evaluating interventionists' criterion on these skills prior to the study. The standards for fidelity during intervention sessions will be reviewed and scoring of fidelity instruments will be discussed such that interventionists are thoroughly familiar with the fidelity standards for the study.

After initial training, caregiver training will be monitored through weekly clinical supervision calls week led by Dr. Hadley and will include all interventionists at both clinical sites. In addition, data on caregiver-training fidelity on each aspect of the TCMR model will be monitored monthly throughout the intervention (during approximately 20% of randomly selected sessions; interventionists will be blinded to sessions selected for fidelity rating). Fidelity data will be used formatively to provide feedback to the interventionists such that any significant deviations or problems in implementing the caregiver-training (fidelity below 80% on any component of caregiver training) are resolved within two intervention sessions after the occurrence. Fidelity data will be summarized across treatment caregiver-child dyads at the end of the study and analyzed for any differences in caregiver-training implementation by site or interventionist.

5.2 PREPARATION/HANDLING/STORAGE/ACCOUNTABILITY

5.2.1 ACQUISITION AND ACCOUNTABILITY

N/A – not a drug, biologic or device trial

5.2.2 FORMULATION, APPEARANCE, PACKAGING, AND LABELING

N/A – not a drug, biologic or device trial

5.2.3 PRODUCT STORAGE AND STABILITY

N/A – not a drug, biologic or device trial

5.2.4 PREPARATION

N/A – not a drug, biologic or device trial

5.3 MEASURES TO MINIMIZE BIAS: RANDOMIZATION AND BLINDING

Because recruitment is continuous, caregiver-child dyads will be randomized following initial screening and determination of biological sex. Dyads will be randomly assigned to EMT-SF or control group by biological sex (M/F) for a total of three strata. The randomization sequence will be generated by a senior statistical analyst based on predefined stratification factors (biological sex) prior to the start of the study. We will use the Psychometrics and Biostatistics Core in the Department of Medical Social Sciences at Northwestern for randomization and data management. The randomization sequence will be uploaded directly to REDCap and all staff members, including the Principal Investigators, will remain blind to the randomization distribution. Once the sequence is uploaded to REDCap, it cannot be accessed or manipulated by any study members. Children will be randomized equally to treatment or business-as-usual (BAU) control groups. Once the screening data for each participant have been entered into the Internet-based system (REDCap), randomization will occur through REDCap by Jeff Grauzer, the NU Statistician who does not interact with any research participants. Specific features of REDCap will be enabled so that randomization is only possible when the specific strata are entered. Further, calculated fields will be utilized to reduce the likelihood that a stratum would be entered incorrectly. Instead, study members will be prompted to enter the number that is automatically generated in the calculated field above to ensure accurate stratification. Similarly, once a participant is randomly assigned, the assignment cannot be manipulated and the ID number cannot be reassigned.

Interventionists and caregivers cannot be naïve to the experimental condition due to the nature of the intervention. All staff members involved in the collection, scoring, and coding of outcomes will be naïve to experimental condition to reduce bias when assessing the child's skills. Staff members involved in coding outcome data will be located at UIUC, which does not provide intervention to any children and families; as such it will be easier for these staff to remain naïve to the experimental condition, given that they will not have day-to-day interactions with interventionists.

Transcribers and coders will be naïve to the experimental condition to reduce bias when coding and scoring the outcome measures. Several steps will be taken to help ensure that study staff

remain naïve to experimental condition. All forms and logs that reveal experimental condition will be hidden on REDCap from study staff members who are naïve to condition (assessors, coders). Similarly, documents saved on the lab server will be separated by assessment data and condition, and appropriate permissions will be given to study staff members. Additionally, a lab onboarding training outlines study compliance, which includes instructions to refrain from discussion or speculation of participant condition. Study staff are also trained that if they accidentally see a document or overhear a conversation that reveals condition, to alert the site principal investigator so that the appropriate steps can be taken.

5.4 STUDY INTERVENTION COMPLIANCE

We will collect the number of intervention sessions that caregivers attend and the length of each intervention session. In addition, we will calculate the therapist fidelity and fidelity of caregiver use of the EMT-SF intervention strategies during each phase and across the 18-month period of the intervention, as described above.

Concomitant Therapy

Caregivers will report frequency, type, and duration of community-based services every 3 months, at all testing time points (30, 33, 36, 39, 42, 45, 49 months of age).

5.4.1 RESCUE MEDICINE

N/A – not a drug, biologic or device trial

6 STUDY INTERVENTION DISCONTINUATION AND PARTICIPANT DISCONTINUATION/WITHDRAWAL

6.1 DISCONTINUATION OF STUDY INTERVENTION

During any testing session, if the child cries for more than 5 minutes, project personnel will suggest that they take a break. If after taking a break, the child continues to cry for more than 5 minutes, project personnel will talk with the caregiver about whether the caregiver wants to continue or if she wants to bring the child back on another day. If the child cries for more than 5 minutes during the second session, project personnel will discuss with the caregiver whether she wants the child to continue to participate in the screening portion of the study. Once a child is in the study, if the child cries for more than 5 minutes during any session, the caregiver will be asked if she would like to end the session. If the child cries for three consecutive sessions, a meeting will be held between project personnel and the caregiver to discuss how to best support continued participation in the study. This has not occurred in any of our previous studies of caregiver-mediated interventions. It is most likely that the child will be dropped (discontinued) only if discontinuation is requested by the caregiver.

6.2 PARTICIPANT DISCONTINUATION/WITHDRAWAL FROM THE STUDY

Participants may withdraw voluntarily from the study or discontinue the study intervention at any time. If a participant is considering withdrawing, study personnel will discuss with the family to

see if there's anything in particular that is influencing their decision and if anything can be done to reduce study burden.

Participants will be withdrawn from the research without their consent if they have missed more than three intervention or assessment sessions without contacting study personnel or if they miss four intervention sessions in a row and cannot reschedule at least two sessions. When a participant withdraws from the study, their previously completed study data will be used in all analyses unless the participant provides a written request for their data to be destroyed. In this case, all participant data will be destroyed. Participants may continue to participate in the study if they partially withdraw from procedures (i.e., miss an intervention session). If a participant is withdrawn, they will be asked to complete the 49-month assessment, given this is an intent-to-treat trial.

6.3 LOST TO FOLLOW-UP

There is no follow-up, see section 5.5 above for strategies to support retention and prevention attrition.

7 STUDY ASSESSMENTS AND PROCEDURES

7.1 EFFICACY ASSESSMENTS

All assessments will be administered virtually, with the exception of assessments at T49. For assessments administered virtually, a trained staff member will coach the caregiver on completing the assessment according to the assessment protocol. All assessments will be administered by trained study staff according to the standardized assessment manual as well as the standard operating procedures of the clinical trial.

Primary

1. *Clinical Evaluation of Language Fundamentals-Preschool, Second Edition (CELF-P2)*

The CELF-P2 will be administered at 49 months of age and will take 10-15 minutes to administer. The CELF- P2 measures a broad range of language skills for preschool children by assessing for aspects of semantics, morphology, and syntax. Testing includes verbal responses to stimulus pictures. Subtests include Concepts and Following Directions, Word Structure, Expressive Vocabulary, Recalling Sentences, Sentence Structure, and Basic Concepts. The Core Language Score, made up of three "core" subtests (i.e., Sentence Structure, Word Structure, Expressive Vocabulary) will be the primary composite measure because it is most discriminating and clinically sensitive in identifying DLD.

2. *Language Sample*

Language samples will be collected via an animated picture book at 3-month intervals at 30, 33, 36, 39, 42, 45, and 49 months of age. It will take 20-30 minutes to administer. The assessor will coach the caregiver on following a standardized protocol for eliciting language for 24 sets of pictures. The language sample will be presented to the child and caregiver on an iPad for standardization across participants. Language samples will be transcribed and coded for the following grammatical elements: (a) number of different

verbs, (b) number of different subjects, (c) number of different subject-verb combinations, and (d) clauses per utterance. In addition, transcripts will be analyzed using Systematic Analysis of Language (SALT) for total number of different root words.

Secondary

1. *Peabody Picture Vocabulary Test – Fifth Edition (PPVT-5)*

The PPVT-5 will be administered to the child at 36 and 49 months of age and will take 10-15 minutes to administer. The PPVT-5 will be administered virtually when the child is 36 months of age. The PPVT-5 measures a broad range of receptive vocabulary levels, from preschool through adults. The items sample words that represent 20 content areas and parts of speech across all levels of difficulty. For each item, the examiner says a word and the examinee responds by selecting the picture that best illustrates the word's meaning.

2. *Expressive Vocabulary Test – Third Edition (EVT-3)*

The EVT-3 will be administered at 36 and 49 months of age and will take 10-20 minutes to administer. The EVT-3 will be administered virtually when the child is 36 months of age. The EVT-3 measures expressive vocabulary and word retrieval. The test is individually administered and quickly assesses expressive vocabulary without requiring reading or writing. Additionally, the test comes with two parallel forms, ideal for pre- and post-testing.

3. *Structured Photographic Expressive Language Test-Preschool – Second Edition (SPELT-P2)*

The SPELT-2 will be administered at 42 and 49 months of age and will take 15-20 minutes to administer. The SPELT-2 will be administered virtually when the child is 36 months of age. The SPELT- P2 assesses broader aspects of morphology, simple syntax, and complex syntax with 44 photographs and corresponding verbal prompts. Up to three verbal prompts may be provided if the child does not attempt the target structure. The SPELT-P2 also demonstrates excellent sensitivity and specificity for accurate identification of SLI in the preschool years, with an empirically derived standard score cut-off of 87.

4. *Test of Early Grammatical Impairment (TEGI) – Third Person Singular and Past Tense Probes*

The TEGI will be administered at 42 and 49 months. The TEGI will be administered virtually when the child is 36 months of age. The phonological probe will be administered first to document children's ability to produce /t, d, s, z/ in non-morphophonemic contexts (e.g., *foot, bed, horse, cheese*). Two TEGI probes will be used: Third Person Singular and Past Tense, which will take a total of 10 minutes to administer. The TEGI provides multiple opportunities to estimate children's progress towards mastery for each structure. Criterion-referenced cut-off scores are provided along with associated levels of sensitivity and specificity. Preschool children who fall below the criterion of 60% correct

use for these two probes combined are unlikely to master T/A by kindergarten entry. Thus, the *TEGI* can be used to identify deficits in a well-established clinical marker of SLI.

5. *Renfrew Bus Story – North American Edition (RBS-NA)*

The Renfrew Bus Story will be administered at 49 months of age and will take approximately 10 minutes to administer. The Renfrew Bus Story is a short screening assessment of receptive and expressive oral language for young children. The assessment incorporates vocabulary and pictures for children to familiarize themselves with elements of the story. Children are then asked to retell the story. This task requires integration of many skills, including coordination of auditory and visual input, attention, listening, comprehension, memory, sentence formulation, and narrative schema understanding. Narrative skills will be measured by the information score (vocabulary/key words) collected during the assessment.

6. *Child Behavior Checklist (CBCL)/1 ½-5*

The CBCL/1 ½-5 will be collected at 30 and 49 months of age and will take 15 minutes to complete. The CBCL/1 ½-5 is a caregiver report form of 99 items used to identify problem behaviors in children between 18 to 60 months of age. Syndrome scales include Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problem, Attention Problems, and Aggressive Problems. Scores include internalizing, externalizing, and total problem scale, and a stress problem scale.

7. *Focus on the Outcomes of Children Under Six (FOCUS)*

The FOCUS will be collected at 30, 33, 36, 39, 42, 45, and 49 months of age and will take 10 minutes to complete. The FOCUS is a clinical tool designed to evaluate change in communicative-participation in preschool children. The FOCUS is a valid, reliable, responsive measure of treatment outcomes that captures 'real world' changes following speech and language treatment.

7.2 SAFETY AND OTHER ASSESSMENTS

Screening

1. *Preschool Language Scale, Fifth Edition (PLS-5)*

The PLS-5 is administered when the child is 30 months of age and will take 30 minutes to administer. The PLS-5 will be administered virtually. The PLS-5 is a play-based assessment that provides comprehensive language information from birth through age 7, including auditory comprehension and expressive communication. Sensitivity for the Total Language score is .83 and specificity is .80.

2. *Mullen Scales of Early Learning – Visual Reception subscale (MSEL-VR)*

The MSEL-VR is administered when the child is 30 months of age and will take 15 minutes to administer. The MSEL-VR will be administered virtually. The MSEL-VR is an assessment designed to measure visual-perceptual skills and nonverbal cognitive abilities in infants and preschoolers. Children complete a variety of matching, sorting, and nesting cups.

3. *Telemedicine-based ASD Evaluation Tool for Toddlers and Young Children (TELE-ASD-PEDS) and Autism Diagnostic Interview – Revised (ADI-R)*

The TELE-ASD-PEDS is administered when the child is 30 months of age and will take 15-20 minutes to administer. The TELE-ASD-PEDS is a telehealth screening tool for autism in children under 36 months of age. The assessor coaches the caregiver on several tasks to complete with their child. The assessment is scored via a 7-item Likert rating form addressing socially directed speech and sounds, eye contact, vocalizations, play, body movements, gestures and sensory exploration. The Autism Diagnostic Interview (ADI-R) will only be conducted if the screening clinicians have clinical concerns about the absence of autism after the child passed the TELE-ASD-PEDS. Parents will report child's autism symptoms in this 90 minute parent interview. This is administered with the child is 30 months of age.

Descriptive

1. *Demographic form*

The Demographic form will be collected from the caregiver every 3 months and will take 10 minutes to complete. This caregiver report measure captures basic information about the family, home structure, and medical history. The form will include key variables, such as maternal education level and positive family history of language impairment.

2. *Community Services form*

The Community Services form will be collected from the caregiver when the child is 30, 33, 36, 39, 42, 45, and 49 months of age and will take 10 minutes to complete. The Community Services form is a caregiver report measure that captures frequency, type, and duration of community services, including additional intervention services that the child is receiving for the duration of the research study.

3. *Autism Screening Survey*

The Autism Screening will be collected from the caregiver when the child is 30 months of age and will take 10 minutes to complete. The Autism Screening Survey consists of 18 caregiver report questions regarding symptoms of autism. This survey includes questions from the CSBS DP Infant-Toddler Checklist, Modified Checklist for Autism in Toddlers (MCHAT) and the Early Signs of Autism Checklist (ESAC). This survey will be used to inform clinical judgement before completing the TELE-ASD-PEDS.

4. *MacArthur-Bates Communication and Development Inventories: Words and Sentences (MCDI-WS)*

The MCDI-WS will be collected from the caregiver when the child is 30 months of age and will take 20 minutes to complete. The MCDI-WS is a caregiver report measure used with children ages 16–30 months. In the first part of each form, caregivers document the child's production and use of hundreds of words divided into semantic categories. The second part analyzes the early phases of grammar, including the child's understanding of word forms and the complexity of the child's multi-word utterances. Caregivers identify the words the child has understood or used and provide written examples of the child's three longest sentences.

5. *Parent Self-Efficacy and Parent Knowledge of Language Survey*

This survey will be completed by caregivers at 30, 36, 42, and 49 months to assess the caregivers' self-efficacy, buy-in, and satisfaction with intervention.

Moderators

1. *Multidimensional Assessment of Preschool Disruptive Behaviors (MAP-DB)*

The MAP-DB will be collected at 30, 33, 36, 39, 42, 45, and 49 months of age and will take 10 minutes to complete. The MAP-DB is a caregiver report form that measures disruptive behavior in early childhood. The MAP-DB measure four dimensions of disruptive behavior, including temper loss, noncompliance, aggressive, and low concern for others.

Mediators

1. *Caregiver-Child Interaction (CCX)*

The CCX will be administered at 30, 33, 36, 39, 42, 45, and 49 months of age. During the caregiver-child interaction, the caregiver and the child will read a standard set of books (5 minutes), play with a standard set of toys (20 minutes) and have a snack with their child (5 minute) and will be instructed to play and communicate as they normally would. Videos of the CCX will be transcribed and coded for the percentage of utterances that contain EMT-SF support strategies. Transcripts will also be analyzed using SALT to measure caregiver NDW, subject diversity, verb diversity, and sentence diversity (i.e., different subjects, verbs, and subject-verb combinations in declarative sentences). Child utterances in the CCX will be transcribed and coded for number of different words, target level utterances, complex sentences, and decontextualized language.

Subgroup

1. *Goldman-Fristoe Test of Articulation, Third Edition (GFTA-3)*

The GFTA-3 will be administered at 49 months of age and will take 20 minutes to administer. The GFTA-3 measures phonological skills by assessing an individual's articulation of the consonant and consonant cluster sounds of Standard American English. It provides information about an individual's speech sound ability by sampling both spontaneous and imitative sound production in single words and connected speech. GFTA-3 provides age-based normative scores separately for females and males for the Sounds-in-Words and Sounds-in-Sentences tests. Intelligibility is reported as a percentage score, and Stimulability information is reported in table format.

2. *Peabody Picture Vocabulary Test, Fourth Edition (PPVT-5) for adults*

The PPVT-5 will also be administered to the primary caregiver when the child is 49 months of age and will take 10-15 minutes to administer. The PPVT-5 measures a broad range of receptive vocabulary levels, from preschool through adults. The items broadly sample words that represent 20 content areas and parts of speech across all levels of difficulty. For each item, the examiner says a word and the examinee responds by selecting the picture that best illustrates the word's meaning.

7.3 ADVERSE EVENTS AND SERIOUS ADVERSE EVENTS

7.3.1 DEFINITION OF ADVERSE EVENTS (AE)

An adverse event is any abnormal, untoward or unfavorable change in health in a human subject, including any abnormal exam or test result, symptom, or disease, occurring during or in close proximity to the end of the subject's participation in the research, whether or not considered related to the research intervention. Adverse events in this study may include:

- a. caregiver psychological stress (caregiver crying for more than 2 consecutive minutes or results in a visit to a mental-health professional),
- b. child distress (child crying for more than 5 consecutive minutes that caregiver reports as atypical)
- c. child injury (falling) that results in crying for more than 2 consecutive minutes after being consoled by caregiver or results in a visit to a medical professional

7.3.2 DEFINITION OF SERIOUS ADVERSE EVENTS (SAE)

Serious Adverse Event (SAE): Any AE (unfavorable change in health) that:

- results in death,
- is life-threatening,
- requires hospitalization or prolongation of hospitalization, or
- results in persistent or significant disability or incapacity
- results in a congenital anomaly/birth defect.
- other important medical events, based upon appropriate medical judgment, may also be considered SAEs if a trial participant's health is at risk and intervention is required to prevent an outcome mentioned.

7.3.3 CLASSIFICATION OF AN ADVERSE EVENT

7.3.3.1 SEVERITY OF EVENT

Severity: The intensity of each AE is defined as follows.

- a. Mild (Grade 1): an event that is easily tolerated by the participant, causing minimal discomfort and not interfering with everyday activities.
- b. Moderate (Grade 2): An event that is sufficiently discomforting to interfere with normal everyday activities.
- c. Severe (Grade 3): An event that prevents normal everyday activities.
- d. Potentially life-threatening (Grade 4): An event that could cause death.
- e. Death (Grade 5): An event that caused death.

7.3.3.2 RELATIONSHIP TO STUDY INTERVENTION

Relationship to Intervention, adverse event attribution categories:

- a. Unrelated: The AE is clearly not related to the intervention
- b. Unlikely: The AE is doubtfully related to the intervention
- c. Possibly: The AE may be related to the intervention
- d. Probably: The AE is likely related to the intervention
- e. Definitely: The AE is clearly related to the intervention

7.3.3.3 EXPECTEDNESS

Expectedness:

- a. Expected AEs: The nature, severity, degree of incidence, specificity, and mechanism is consistent with the components of the intervention.
- b. Unexpected AEs: Previously unobserved or documented, the nature, severity, degree of incidence, specificity, and mechanism is not consistent with the components of the intervention or study procedures.

7.3.4 TIME PERIOD AND FREQUENCY FOR EVENT ASSESSMENT AND FOLLOW-UP

Adverse events will be collected by caregiver and clinician report. At the beginning of each assessment visit after baseline (33, 36, 39, 42, 45 and 49 months) and intervention session, the assessor or interventionist will ask the caregiver, "Have there been any significant problems that may have impacted your child's development since your last visit?" Assessors will transcribe the caregivers' responses to this open-ended question. In addition, caregivers will be asked specifically for the presence of the aforementioned predefined adverse events as part of the demographic form collected at 33, 36, 39, 42, 45 and 49 months.

Within 8 hours after the completion of the session, assessors and interventionists will complete a session log. They will enter the caregivers' responses into a free-form box. They will also determine if the caregiver reported an adverse event or if they observed an adverse event during the session. Adverse events are predefined before the start of the study (as described

above) and research staff are prompted to classify: (a) the type of adverse event (adverse or serious adverse event), (b) the severity of the event (grade 1 through grade 5), (c) the relatedness of the event to study procedures (unrelated to definitely related), (d) the expectedness of the event (expected or unexpected), and (e) a more detailed description of the event. These session logs are monitored weekly by the project manager and discussed with the PI. We will follow up on all grade 2 to grade 5 adverse events by calling or emailing the caregiver within 24 hours and followed until stabilization or resolution up to 1 month after the participant discontinues from the study.

7.3.5 ADVERSE EVENT REPORTING

Any adverse events that are possibly, probably, or definitely related to the study procedures will be reported to Northwestern University's IRB within 24 hours. A detailed written report will also be submitted within 5 days to Northwestern University's IRB. Additionally, these adverse events will be Northwestern University's IRB on a yearly basis during the Continuing Review submission. Other AEs definitely not related to the study will not be reported to the IRB.

7.3.6 SERIOUS ADVERSE EVENT REPORTING

Any serious adverse events (SAEs) that are possibly, probably, or definitely related to the study procedures will be reported to the NIDCD Medical Officer and Northwestern University's IRB within 24 hours. In addition, a detailed written report will be submitted within 5 days to NIDCD and the Northwestern University's IRB. An unanticipated death that is not related to study activities will be reported to IRB on a yearly basis. **All deaths and life-threatening serious adverse events** will be reported by email to the NIDCD Medical Officer within 24 hours regardless of relatedness. Other SAEs definitely not related to the study will not be reported to the IRB.

7.4 UNANTICIPATED PROBLEMS

7.4.1 DEFINITION OF UNANTICIPATED PROBLEMS (UP)

The phrase "unanticipated problems involving risks to subjects or others" is found in the HHS regulations at 45 CFR part 46. OHRP considers *unanticipated problems*, in general, to include any incident, experience, or outcome that meets **all** of the following criteria:

1. unexpected (in terms of nature, severity, or frequency) given (a) the research procedures that are described in the protocol-related documents, such as the IRB-approved research protocol and informed consent document; and (b) the characteristics of the subject population being studied;
2. related or possibly related to participation in the research (in this guidance document, *possibly related* means there is a reasonable possibility that the incident, experience, or outcome may have been caused by the procedures involved in the research); and
3. suggests that the research places subjects or others at a greater risk of harm (including physical, psychological, economic, or social harm) than was previously known or recognized.

7.4.2 UNANTICIPATED PROBLEM REPORTING

If unanticipated problems require a change to the research protocol, the unanticipated problem will be reported within 2 weeks to the NIDCD Medical Officer. Prior to changing any study protocols, we will notify Northwestern University's IRB and the NIDCD Medical Officer.

7.4.3 REPORTING UNANTICIPATED PROBLEMS TO PARTICIPANTS

If an unanticipated problem requires a change to the research protocol or consent and current participants are affected, the study team will notify currently enrolled participants of changes and re-consent them if necessary.

8 STATISTICAL CONSIDERATIONS

8.1 SAMPLE SIZE DETERMINATION

Power analyses were carried out using R's pwr library, OptimalDesign, and verified by writing simulation studies that more closely represented the longitudinal design of our study. We completed power analyses on our three primary aims, all assuming a Type I error rate (alpha) of .05 and a net sample size of $n=92$, or 15% attrition from our total sample of 108. Additional details about planned analyses are included in Table 7.

Our first primary outcome looks at intervention effects on CELF-P2 core language score at 49 months. We include a baseline covariate ($r=.70$, $R^2=.49$) to improve power and reflect the strong effect of baseline language on all analyses, as recommended in other power estimation methods for clinical trials (Glueck & Muller, 2003; Hauck, Anderson, & Marcus, 1998; Hernandez, Steyerberg, & Habbema, 2004). We project 80% power for an effect size of $d=.39$ or larger, and 92.7% power to detect the mean effect size of $d=0.45$ found in Roberts and Kaiser (2011) meta-analysis of caregiver intervention effects on overall language scores. An effect size of .39 corresponds to a difference of 6 standard score points on the CELF-P2, which is a clinically significant difference in the language skills, given that an increase in standard scores would reflect a change in growth rate above and beyond the average rate of growth for typical peers. A standard score of 6 points is also larger than the standard error of measurement for this assessment.

Our second and third primary aims deal with intervention effects on slope of language parameters from latent growth curve models. Intercept terms from these models (centered at baseline assessment) provide a comparable baseline covariate as those found in the previous univariate power calculation. Latent growth curve simulations were used to estimate power under varying assumptions to make our estimates more robust, specifically the rate at which children reach the final effect size. As we do not have a priori estimates of intra-individual variability in rates of learning and development, we used the same effect size in our simpler analyses ($d=.45$) and used this effect as difference between groups at 49 months. Thus, an effect size of $d=.45$ over 18 months yields an effect of $d=.30$ per year.

Under the most conservative assumptions (slowest growth, $n=92$) and an assumed correlation between level and change of $r=.50$, we estimate 80% power to determine effect sizes of $d=.48$ or larger. Under more relaxed assumptions (growth occurs largely during treatment, $n=92$), we find 80% power for effect sizes of $d=.44$ or larger. Under the most favorable conditions (growth occurs during treatment, $n=108$ at baseline, 15% attrition during first year), we find 80% power for effect sizes of $d=.41$ or larger). Finally, applying a Bonferroni correction and using $p=.01667$ as a criterion value yields 80% power under the most conservative conditions for $d=.57$, and 90% for $d=.63$. Latent growth curve models estimated via either structural equation models or multi-level models contain a sufficient number of time points and sufficient sample size (time points * sample size = 700) to assess longitudinal change, test for nonlinear change, and use that change to predict outcomes, based on published simulation studies. Taken together, our longitudinal design allows sufficient power for primary hypotheses and our secondary and exploratory analyses.

8.2 POPULATIONS FOR ANALYSES

Intent to treat (ITT) analyses will be performed using R and SAS with all randomized participants.

8.3 STATISTICAL ANALYSES

General statistical approach. All statistical models will be fit within a structural equation modeling (SEM) paradigm. Our proposed models vary in complexity, from simple linear regressions of main intervention effects to longitudinal mediation analyses. Providing a unified modeling framework for all models will allow for comparability of related analyses due to identical model assumptions. While SEM was traditionally developed as covariance structure analysis, modern or extended SEM (xSEM) can flexibly handle both linear and nonlinear relationships between sets of continuous and categorical raw data. This flexibility allows us to specify SEMs that are equivalent to alternatives ranging from linear regression to mixed-effect models to complicated mediation analyses, all using the same state-of-the-art optimization and data handling methods. For example, vocabulary skills are measured both: (a) continuously during intervention (via a language sample) and (b) after intervention (via the Peabody Picture Vocabulary Test; PPVT-5 and the Expressive Vocabulary Test, EVT-3). As such, analysis of the effects of intervention on vocabulary skills features two similar analyses: a latent growth curve of the language samples and a linear regression for vocabulary scores. These two analyses answer similar questions (e.g., what is the direct effect of intervention on child vocabulary outcomes) that differ based on the data on which they depend. These two modeling frameworks vary in their complexity and their modeling assumptions, specifically that latent growth curves estimated via SEM have more robust missing data handling than conventional regression software that requires complete cases. Estimating all models as SEMs means that all models retain the same underlying full sample sizes and assumptions, so that disparate results can be more directly attributed to constructs. In this example, stronger effects on vocabulary skills from language samples would reflect more proximal effects of intervention rather than a weakness of conventional regression methods and their missing data assumptions.

Multicollinearity. This proposal includes a large set of linguistic variables, some of which will be included in the same analyses. Our data reduction is one part of our solution to multicollinearity, such that latent constructs that account for shared variation between

comparable predictors will be used when possible. In other cases, we will control for multiple collinear variables in the same analysis, specifically when using multiple coded variables from naturalistic caregiver-child interactions (CCX). In those cases, we will use alternative decomposition and transformation methods. Principle components analysis (PCA) can transform a set of collinear predictors into a set of uncorrelated variables representing variance in the original data. For example, a simple PCA of one caregiver and one child variable would yield two components: one indicating shared family level behavior and the other representing scaled differences between caregiver and child. We will further explore nonlinear transformations to make sure nonlinear associations and distributional assumptions are accounted for, including both simple (e.g., polynomial or logarithmic transformations to improve distributions) and complex transformations. This set of methods integrates well within our SEM framework and multivariate longitudinal design.

Missing data. Given this is an intent-to-treat trial, all children randomized will be included in the analyses of primary outcomes even if data are missing. Our preliminary studies showed high retention rates, but missing data occurs in all longitudinal studies. All models will be estimated using full information maximum likelihood (FIML), which uses all available data and yields unbiased parameter estimates under the most common missingness mechanisms. We will further use our set of predictors and covariates to predict missingness, and include those variables as covariates to conform to FIML assumptions. For non-ignorable missing data, we will conduct sensitivity analyses using non-ignorable pattern-mixture and selection models to investigate the robustness of our conclusions across the different models for missing data.

Covariates. Controlling for covariates is an important part of any analysis. While randomization and stratification should remove effects of these variables, additional statistical control curtails potential unhappy randomizations and other threats to internal validity. As such, we will examine differences between experimental conditions for the following variables: nonverbal cognitive skills, speech sound skills, SES, other spoken languages in the home, and amount of community-based early intervention therapy services. Should a difference exist at any point in the study, we will include that variable in all statistical analyses.

Multiple testing. Our proposed analyses and rich multivariate data collection will allow us to test hypotheses in multiple ways. Our focus on SEM will allow for complex hypotheses to be tested with a single model, obviating part of the need for multiple testing corrections. However, we will further employ Benjamini-Hochberg corrections (Mangold, 1998) as needed to correct for multiple testing. This procedure is a more sophisticated alternative to traditional Bonferroni corrections that better account for multiple simultaneous significant parameters.

Table 7 provides a summary of research questions and specific analyses; descriptions of specific analyses follow in the text below.

Table 7. Research Questions and Specific Analyses

Analysis	Research Question	Hypotheses	Outcome	Outcome Time Point	Analysis
1.1	What are the long-term effects of the intervention on overall	Children in the EMT-SF group will have better overall language scores	CELF-P2 Core Language Score at 49 months	49 months	Linear regression with Preschool Language Scale Total Standard

Analysis	Research Question	Hypotheses	Outcome	Outcome Time Point	Analysis
	language skills at 49 months of age?	than children in the control group.			Score at 30 months (T0) as a covariate.
1.2	What are the long-term effects of the intervention on growth in vocabulary skills during intervention (from 30 to 49 months)?	Children in the EMT-SF group will have better vocabulary growth than children in the control group.	Slope of total number of different spoken word (NDW) roots from the Language Sample measured every 3 months from 30 to 49 months	Intercept at 49 months	Multi-level growth curve (30, 33, 36, 39, 42, 45, 49 months)
1.3	What are the long-term effects of the intervention on growth in grammar skills during intervention (from 30 to 49 months)?	Children in the EMT-SF group will have better grammar growth than children in the control group.	Slope of a latent variable: - Number of different verbs - Number of different subjects - Clauses per utterance from the Language Sample measured every 3 months from 30 to 49 months	Intercept at 49 months	Latent factor multi-level growth curve model (30, 33, 36, 39, 42, 45, 49 months)
2.1	What are the short-term effects of the phase 1 vocabulary intervention on child vocabulary?	Children in the EMT-SF group will have better short-term vocabulary outcomes than children in the control group.	Latent Variable: - PPVT standard score - EVT standard score At 36 months	36 months	Linear regression with NDW from the language sample at 30 months (T0) as a covariate
2.2	What are the short-term effects of the phase 2 grammar intervention on child grammar?	Children in the EMT-SF group will have better short-term grammar outcomes than children in the control group.	Latent Variable: - SPELT-P2 raw score - TEGI raw score At 42 months	42 months	Linear regression with IPSyn from the language sample at 30 months (T0) as a covariate.
2.3	What are the effects of the intervention on child communication and interaction in real world settings at 49 months of age?	Children in the EMT-SF group will have better functional communication skills than children in the control group.	Focus on the Outcomes of Children Under Six (FOCUS) Raw Score	Intercept at 49 months	Multilevel growth curve (30, 36, 42, 49 months) with FOCUS raw scores at 30 months (T0) as a covariate.
2.4	What are the effects of the intervention on rate of diagnosis of developmental language disorder?	There will be fewer children diagnosed as having a developmental language disorder in the EMT-SF group than in the control group.	Diagnosis of a developmental language disorder based on SPELT-P2 and TEGI raw scores	49 months	Logistic regression with Preschool Language Scale Total Standard Score at 30 months (T0) as a covariate.
2.5	What are the effects of the intervention on decontextualized language (DL)?	Children in the EMT-SF group will use more decontextualized language than children in the control group.	Total number of utterances that contain decontextualized language from a language sample at 49 months	49 months	Linear regression with Preschool Language Scale Total Standard Score at 30 months (T0) as a covariate.
2.6	What are the effects of the intervention on narrative skills?	Children in the EMT-SF group will have better narrative skills than children in the control group.	Information score from the Renfrew Bus Story at 49 months	49 months	Linear regression with Preschool Language Scale Total Standard Score at 30 months (T0) as a covariate.

Analysis	Research Question	Hypotheses	Outcome	Outcome Time Point	Analysis
2.7	What are the effects of the intervention on internalizing behaviors?	Children in the EMT-SF group will have fewer internalizing behaviors than children in the control group.	Child Behavior Checklist Internalizing T score at 49 months	49 months	Linear regression with CBCL at 30 months (T0) as a covariate.
3.1	Do baseline receptive language skills (PLS-AC standard score) moderate intervention outcomes?	Children with better receptive language skills will have a larger intervention effect than children with poorer receptive language skills.	Same as the 3 primary analyses	Same as the 3 primary analyses.	Same as the 3 primary analyses with Preschool Language Scale – Auditory Comprehension Standard Scores at 30 months (T0) and an interaction term between PLS-AC and experimental condition
3.2	Does biological sex moderate intervention outcomes?	Girls will have a larger intervention effect than boys.	Same as the 3 primary analyses	Same as the 3 primary analyses.	Same as the 3 primary analyses with sex and an interaction term between sex and experimental condition.
3.3	Does family history of language impairment moderate intervention outcomes?	Children with a negative family history of language impairment will have a larger intervention effect than children with a positive family history of language impairment.	Same as the 3 primary analyses	Same as the 3 primary analyses.	Same as the 3 primary analyses with positive family history (yes/no by caregiver report) and an interaction term between family history and experimental condition.
3.4	Does presence of disruptive behaviors at baseline moderate intervention outcomes?	Children with fewer disruptive behaviors will have a larger intervention effect than children with more disruptive behaviors.	Same as the 3 primary analyses.	Same as the 3 primary analyses.	Same as the 3 primary analyses with raw score on the Multidimensional Assessment of Preschool Disruptive Behavior (MAP-DB) and an interaction term between MAP-DB and experimental condition
3.5	Does caregiver use of child-directed, developmentally appropriate, target language mediate intervention outcomes?	Caregiver use of child-directed, developmentally appropriate, target language will mediate intervention outcomes.	Same as the 3 primary analyses.	Same as the 3 primary analyses.	Mediation analysis with a number of adult utterances from the caregiver child interaction that include child-directed, developmentally appropriate, target language measured

Analysis	Research Question	Hypotheses	Outcome	Outcome Time Point	Analysis
					every 3 months from 30 to 49 months as the mediator.
3.6	Does early child growth in vocabulary and grammar during intervention mediate outcomes? (see Figure 4)	Both growth in grammar and vocabulary will mediate long-term child language outcomes.	CELF-P2 Core Language Score at 49 months	49 months	Multiple mediation analysis with the following mediators: • Total number of different spoken word (NDW) roots from the language sample measured at 30, 33 and 36 months. • A latent variable: number of different verbs, number of different subjects,, Clauses per utterance from the language sample measured at 36, 39 and 42 months.

Regression for primary and secondary intervention effects. The simplest analyses used in this proposal are multiple regressions, which allow for direct comparisons between intervention and control groups with minimal assumptions and the inclusion of covariates. These regressions will include experimental condition (intervention, control) as one predictor and other relevant covariates. These models can be fit either in traditional omnibus software (e.g., R, SPSS, SAS) or via SEM software to take advantage of FIML when attrition is present, as the primary difference between traditional and SEM manifest variable regression approaches is missing data handling.

Latent Growth Curve Modeling. Latent growth curve models are common tools for assessing both the direction and magnitude of how people change, as well as individual differences in the rates of change. These models hypothesize latent variables that represent between-person differences in both overall level (intercept) and rates of change (slopes). These models are commonly fit as structural equation models, but can be equivalently specified within mixed effect models (also called multilevel modeling and hierarchical linear modeling). Latent growth curves can also: (a) be easily extended into nonlinear change, (b) incorporate intervention and covariate effects, and (c) include cross-lagged regression terms like those used in mediation analyses. While many applications of growth curve models use observed outcomes (e.g., repeated measurement of vocabulary and grammar skills from the language sample), latent growth factors can predict latent variables just as easily under the CUFFS, or curves-of-factors approach.

Moderation. Moderation in clinical trials deals with differential effectiveness of intervention as a function of other independent variables. This is typically accomplished with interaction terms, which mathematically state that the effect of an intervention goes up with increases in specific covariates. We will employ all current standards for regressions using interaction terms,

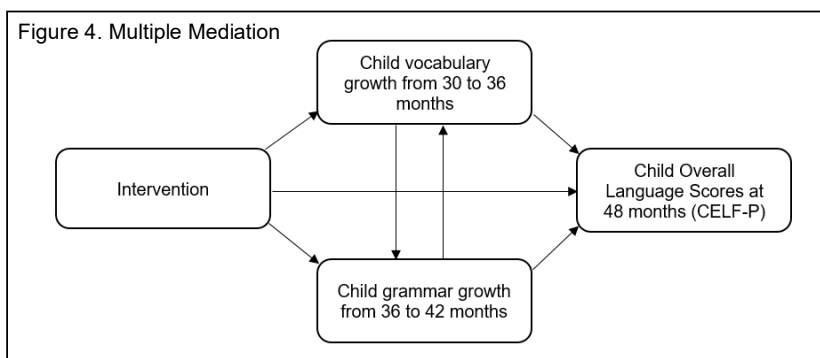
including variable centering, application of variance inflation factors to test for multicollinearity, and residual diagnostics to ensure valid and interpretable estimates.

Multiple-group SEM. Interaction can be further implemented via multiple-group SEM due to the fact that our predictor (intervention status) is a categorical variable. Intervention and control groups are specified as separate groups in a multiple-group SEM framework, and all model parameters can either be constrained to equality across groups to measure main effects or freed to test interactions and moderation. This method allows for both equivalence to extant moderation approaches and flexibility to test for differences in nontraditional parameters, specifically residual variance terms.

Mediation. Exploratory mediation analyses address intermediary mechanisms of the intervention by way of mediation analysis. Mediation analyses consist of a broad set of methods that test the extent to which causal pathways between variables can be explained through effects on a third variable.

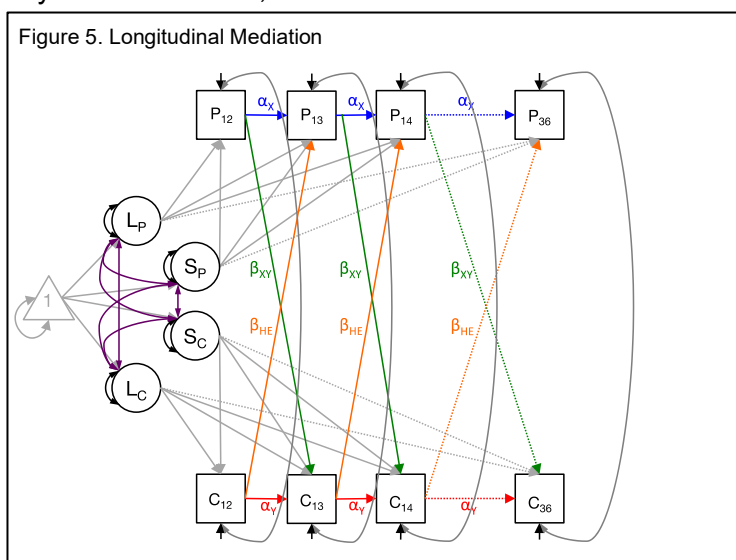
Cross-sectional mediation. Basic cross-sectional mediation analyses typically involve three regressions: the mediator is regressed on the predictor (the 'a' path), and the outcome is regressed on both the mediator (the 'b' path) and the predictor (the 'c' path or total effect of the predictor on the outcome). The total effect of the predictor on the outcome is then split into the indirect effect, which describes what part of the total effect goes through the mediator (calculated as $a*b$), and the direct effect, which describes what part of the total effect is independent of the mediator. While early methods for mediation involved the fitting of separate multiple regression models, more modern methods employ structural equation modeling to fit all model parameters simultaneously. This allows for more rigorous testing of indirect effects, including more accurate standard errors and improved missing data handling.

Figure 4. Multiple Mediation



Longitudinal mediation. An important part of mediational tests is the passage of time, as predictors must precede the variables they cause. As such, mediation models have been extended into longitudinal modeling

(as shown in Figure 5). This model extends the cross-lagged panel design model, such that both the ‘a’ path (from the predictor to the mediator) and the ‘b’ path (from the mediator to the outcome) span one lag (e.g., from age 30 months to 33 months, from 33 months to 36 months, etc.), while the direct effect of the predictor on the outcome (the ‘c’ path) spans two lags (e.g., from age 30 months to 36 months, 33 months to 39 months, etc.). The inclusion of these time lags allows for the use of all measurement occasions simultaneously while also ensuring that the effects required in any mediation analysis show the temporal ordering to test causal direction. As with all other analyses, we will include relevant covariates in all steps of both the cross-sectional and longitudinal mediation models.



Timing of effects and extensions of longitudinal mediation. Longitudinal mediation models provide many benefits over cross-sectional mediation, but a weakness of this approach is that it is tied to the lags inherent to our study design. The intervention predicts growth in child language skills, which we assess every 3 months for longitudinal modeling. However, children may develop either more quickly or more slowly than these intervals, which is not traditionally accounted for in longitudinal mediation analyses.

Modern extensions of longitudinal mediation utilize the benefits of these approaches with a more continuous approach to lags. DeBoeck and Preacher’s (2016) continuous time longitudinal mediation model merges existing longitudinal mediation methods with dynamical systems modeling to estimate not only the size of the mediational paths, but also the optimal lag for each parameter. We hypothesize that intervention effects on caregiver linguistic behavior will occur relatively quickly, while the effects on child outcomes will develop more slowly over time.

Latent Growth Curves and Autoregressive Latent Trajectories. The above mediation models provide ways of testing our primary hypotheses, but also allows us to test secondary questions about the shape and timing of change in all studied constructs. Latent growth curve models specify that change in a measured variable can be modeled via latent variables representing initial level and change, each of which are free to vary across individuals. The Autoregressive Latent Trajectories model (ALT) is the combination of a latent growth curve and a cross-lagged regression model (Figure 5), which is commonly used in longitudinal mediation. Specifically, this model takes a bivariate growth curve model of caregiver and child language behaviors and adds autoregressions (such that caregiver use of communication support strategies at early time points predicts later caregiver use of communication support strategies; see the “alpha” parameters) and cross-regression (such that earlier caregiver use of communication support strategies predicts later child communication skills; see the “beta”

parameters). Use of the ALT model will allow improved model fit and greater accuracy in specifying the longitudinal trends in both caregiver behaviors and child language outcomes.

These growth curves will further allow a merging of the cross-sectional and longitudinal mediation paradigms described above. These approaches will use basic cross-sectional methods with single variables representing the predictor, mediator, and outcome. However, the mediators and outcomes in these models will be latent variables that are estimated within a latent growth curve model.

Multiple Mediation. Cross-sectional mediation approaches can be extended to test for multiple mediators, as the intervention is hypothesized to improve growth in both child vocabulary and child grammar, which in turn explain part of the effects of intervention on overall child language (CELF-P2) at 49 months. As these mediators may be correlated, it is important to include both simultaneously in addition to traditional single mediator models as discussed above. Multiple moderation may be carried out using the same structural equation modeling approach as single mediation, and allows for exploratory analyses of interactions and associations between the pair of mediators (VanderWeele & Vansteelandt, 2014). This will also allow us to differentiate the effects of vocabulary and grammar, or test whether such a distinction is needed.

8.3.1 SAFETY ANALYSES

Interim analyses will be conducted by Dr. Estabrook at each annual review for safety only, not effectiveness. Assuming there are no safety issues, we will follow all children and caregivers for 18-19 months, and analyze the data when collection is completed.

8.3.2 BASELINE DESCRIPTIVE STATISTICS

Age in months, biological sex, maternal education, SES, race/ethnicity, number of children living in the home, no-verbal thinking skills (Mullen visual reception t-scores), speech sounds inventory (PEEPs score), amount of exposure to languages other than English, number of words spoken by the child, baseline values of all outcome measures, use of multi-word utterances, length and frequency of additional speech-language and other educational therapy.

8.3.3 PLANNED INTERIM ANALYSES

Interim data examination for possible early termination are not relevant in this trial. We are providing a service over and above what children and caregivers would typically receive; services are not being withheld or denied.

8.3.4 SUB-GROUP ANALYSES

Subgroup analyses will be conducted for:

1. Children with nonverbal cognitive skills of 35-40 (criteria for DLD) and 41 and above (criteria for SLI) on the Mullen Visual Reception Scale.
2. Children with and without speech sound disorders.
3. Parents with and without language impairment observed via direct testing.

8.3.5 TABULATION OF INDIVIDUAL PARTICIPANT DATA

All participant data will be listed by measure and time point.

9 SUPPORTING DOCUMENTATION AND OPERATIONAL CONSIDERATIONS

9.1 REGULATORY, ETHICAL, AND STUDY OVERSIGHT CONSIDERATIONS

9.1.1 INFORMED CONSENT PROCESS

9.1.1.1 CONSENT/ASSENT AND OTHER INFORMATIONAL DOCUMENTS PROVIDED TO PARTICIPANTS

All consent forms will be approved by a single IRB (Northwestern University's Institutional Review Board), the NIH PO, and the DSMB. Consent forms for Vanderbilt University participants will use the VUMC Human Research Protections Program template and receive separate approval from the Vanderbilt University Institutional Review Board. All informed consent documents will contain the following: (a) an explanation of the purpose of the research, (b) expected duration of participation, (c) a description of research procedures, (d) foreseeable risks, discomforts, and benefits, (e) confidentiality and privacy of records and data, (f) contact information for whom to contact to answer questions about the research process or participant rights, and (g) a statement that participation is voluntary and that refusal will not result in penalty or change in their clinical services. Additional consent elements will also include financial information, alternative treatments, funding source, mandated reporter language, clinical trial information, and optional elements that allow (1) video recordings to be used for training and educational purposes, (2) future contact about participating in research, (3) video recording clips (1-2 minutes) of caregiver/child participation may be shown to future families participating in the Early Intervention Research Group's or KidTalk research studies, (4) data collected in this study may be used for future research questions and analysis and (5) (NU only) if enrolled in study STU00208022, permission for data from the current study to be used in the TODLS study. All consent documents will contain layman's terminology when possible and use language appropriate to an 8th grade reading level. Consent documentation will include the IRB watermark with the date of IRB approval and the expiration date.

9.1.1.2 CONSENT PROCEDURES AND DOCUMENTATION

The participating caregiver will sign an electronic consent form outlining the study characteristics and requirements. Research personnel will review the consent form with all participating caregivers, and caregivers will be given a chance to ask questions about their participation before signing the consent. The caregiver and the research staff will each keep a signed and dated copy of the consent form. Consent forms will be stored electronically on a secure and password-protected server and/or a hard copy will be kept in a locked filing cabinet behind a locked door. Child assent will not be sought due to the child's age. Although permission for child participation is obtained from the caregiver, children will be watched for dissenting behaviors. Caregivers will also remain with or be able to view the child at all times so they can monitor their child.

All informed consent documents will be reviewed by the Northwestern Project manager after they are obtained to check for proper document, including that all consent documents are signed, dated, initialed, and contain the appropriate IRB watermark.

The Institutional Review Board (Committee for Protection of Human Subjects-Behavioral Sciences) at Northwestern University will review this project for approval under the standard review process and follow the conditions under subpart §46.408 of the Code of Federal Regulations. Consent/assent procedures will be followed according to these guidelines as outlined above and determined by Northwestern University's Institutional Review Board.

9.1.2 STUDY DISCONTINUATION AND CLOSURE

This study may be temporarily suspended or prematurely terminated if there is sufficient reasonable cause. Written notification, documenting the reason for study suspension or termination, will be provided by the suspending or terminating party to the NIDCD. If the study is prematurely terminated or suspended, the site principal investigators at Northwestern and Vanderbilt will promptly inform study participants, the Institutional Review Board (IRB), and sponsor and will provide the reason(s) for the termination or suspension. Study participants will be contacted, as applicable, and be informed of changes to study visit schedule.

Circumstances that may warrant termination or suspension include, but are not limited to:

- Determination of unexpected, significant, or unacceptable risk to participants
- Insufficient compliance to protocol requirements
- Data that are not sufficiently complete and/or evaluable
- Determination that the primary endpoint has been met

Study may resume once concerns about safety, protocol compliance, and data quality are addressed, and satisfy the sponsor, and the IRB.

9.1.3 CONFIDENTIALITY AND PRIVACY

The only potential risk to study participants associated with this project is the risk of confidentiality being violated. Protection of participant confidentiality is foremost, and there are safeguards in place to minimize this risk. All study data will be stored in secure locations that are locked or password protected. Access to personal identifying information and study data will be limited to research personnel.

All data, including assessment protocols, surveys, and video records, are stored by a unique identification number assigned to each study participant. All study data are stored using: (a) REDcap (Research Electronic Data Capture), a secure, web-based, HIPAA-compliant system for collecting and managing data, and (b) a secure password-protected server at each research site and also on Northwestern University's server for cross-site monitoring and data analysis. Only study research personnel will have password-protected access to both the databases and the secure server. Two separate databases will be created in REDCap in order to keep personal identifying information and study ID number separate. The first database includes de-identified study data, which is stored by ID number. The link between name/contact information and ID number is stored in a separate second database to which only the PIs at Northwestern and Vanderbilt and research personnel have access.

All hard copies of study documents will only contain the child's unique study identification number. All hard copies will be scanned onto the secure server and uploaded to REDCap. Hard copies and video cameras will be stored in a locked lab office until upload. Following upload, data entry, and verification, the electronic files on the video recorders will be deleted. Research personnel will be responsible for these data management tasks. All hard copies of the informed consent will be kept in a locked filing cabinet behind a locked door that only research personnel can access.

Research staff who have access to study information include: the principal investigators, research assistants, research clinicians, the statistician, and the project manager. We foresee no circumstance in which personal identifiable information would be released to a third party, including the funding agency or representatives of the University's IRB. If any data are shared with a third party, no personal identifiable information will be included and participants will only be identified by their unique identification number.

Should we incidentally discover other developmental or medical concerns, we will refer the child to the appropriate medical professional. Should we incidentally discover signs of suspected child abuse or neglect, we will call the Illinois or Tennessee Department of Children and Family Services 24-hour child-abuse hotline.

Data will be stored for at least 7 years after the completion of the study. Findings will be reported in the aggregate with no individual identification associated with any reported data. Should the results of the project be presented at conferences, seminars, and/or workshops, the identity of the participants will not be disclosed. The video recordings will only be shown if permission was obtained in the informed consent from the specific caregiver.

9.1.4 FUTURE USE OF STORED SPECIMENS AND DATA

For the purpose of this project no specimens were obtained for research and no genetic testing will be conducted. Caregivers can agree or disagree in the consent document that their video recordings be used for teaching and training purposes. Only caregivers that agreed to this optional element of the consent will have their video recordings utilized for future training purposes.

9.1.5 KEY ROLES AND STUDY GOVERNANCE

Principal Investigator

*Megan Roberts, PhD, CCC-SLP
Northwestern University
2240 Campus Drive,
Evanston IL 60201
847-491-3183
megan.y.roberts@northwestern.edu*

Principal Investigator

*Ann Kaiser, PhD
Vanderbilt University
One Magnolia Circle,
Nashville, TN 37203
615-322-8160
ann.kaiser@vanderbilt.edu*

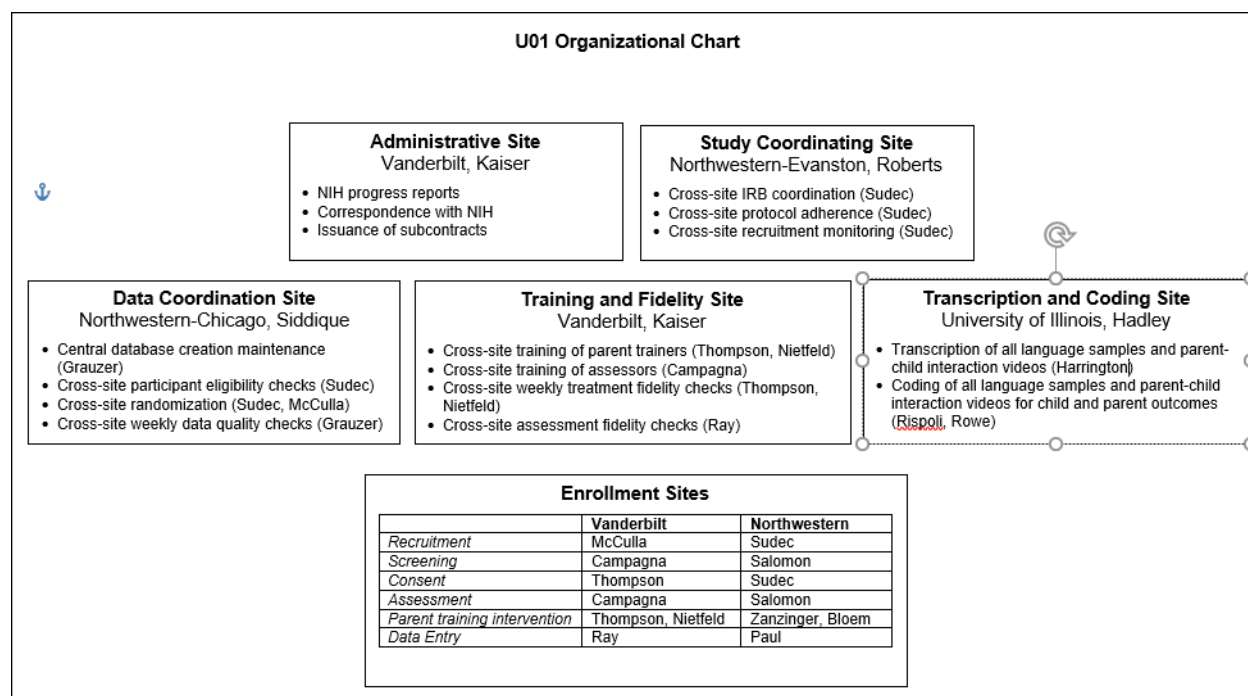
Principal Investigator

*Pamela Hadley, PhD, CCC-SLP
University of Illinois
901 South Sixth Street
Champaign, IL 61820
217-333-1968
phadley@illinois.edu*

The three PIs will form a Steering Committee that will manage the oversight and coordination of project management, research administration, publications and data sharing, and integration of all resources needed for the project.

Dr. Kaiser will lead the Vanderbilt site and implement the treatment protocol for half the participants. Dr. Kaiser will assume overall responsibility for the implementation and fidelity of the treatment and control group protocols across sites. Dr. Roberts will lead the Northwestern site, which will implement the treatment and control protocols for half of the participants. Dr. Roberts will assume overall responsibility for IRB management, data management, and statistical analysis of the data with Dr. Estabrook. Dr. Hadley will lead the University of Illinois site and lead the linguistic data coding and analysis with assistance from Dr. Rispoli and consultant Dr. Rowe. Each PI will manage the research staff at their respective sites. An overview of the organization of the project is shown in Figure 6 below.

Figure 6. Organizational Chart



The PIs will communicate weekly, either by videoconferencing, phone, or email, to discuss experimental design, data analysis, and all administrative responsibilities. If a potential conflict develops, the PIs shall meet and attempt to resolve the dispute. If they fail to resolve the dispute, the disagreement shall be referred to an arbitration committee consisting of one impartial senior executive from each PI's institution and a third impartial senior executive mutually agreed upon by both PIs. No members of the arbitration committee will be directly involved in the research grant or disagreement.

If one of the PIs moves to a new institution, attempts will be made to transfer the relevant portion of the grant to the new institution. In the event that a PI cannot carry out her duties, a new PI will be recruited as a replacement, subject to the approval of the Steering Committee and the Institution.

9.1.6 SAFETY OVERSIGHT

A Data and Safety Monitoring Board will monitor the project at the three sites and will report to NIDCD and the IRBs at each site. The primary role of the DSMB is to monitor the safety of participants in both the intervention and the control condition, validity and integrity of the data, and conduct of the study. The DSMB will be composed of four members, none of whom are directly involved in the project. The DSMB will consist of four experts knowledgeable in several areas relevant to this study including early intervention, caregiver training, language development in preschool children, advanced statistical analysis, and randomized clinical trials.

The DSMB will meet first to approve the protocol and subsequently conduct annual reviews to determine whether patient safety has been adequately safeguarded and enrollment goals have been met. The DSMB will be kept apprised of serious adverse events related to the study intervention annually. The PO will be notified monthly and project personnel will discuss weekly. The statistician will prepare annual reports summarizing the information required for each evaluation, including data quality, completeness, and timelines, adverse events, performance of each site, compliance with recruitment and retention goals, adherence to the protocol, and any other factors which might affect the study outcome or compromise confidentiality of the trial data.

9.1.7 CLINICAL MONITORING

All interventionists will receive training on all intervention components; all interventionists will achieve a fidelity rating of 80% or higher on three consecutive sessions prior to delivering the intervention (both the child-directed intervention and the caregiver training protocol) with research participants. Furthermore, monitoring and maintaining intervention fidelity will occur at two levels: (a) the interventionist's use of intervention strategies during her interactions with the child, and (b) the delivery of caregiver training sessions by the interventionist. Every session will be video recorded to monitor fidelity at each level using a checklist that includes all aspects of the intervention session (modeling the strategies with the caregiver, coaching the caregiver to use the strategies). Fidelity of intervention administration will be assessed on a random sample of 20% of the intervention sessions, distributed across families, sites, phases of the intervention, and interventionists. In addition, we will complete reliability for 20% of the sessions rated for fidelity. Fidelity assessments will be completed before the next intervention session to provide timely feedback to the interventionist. Fidelity will be completed by the Vanderbilt University team and Northwestern University team and overseen by Dr. Kaiser. Timely completion of fidelity assessments will be monitored by Northwestern University under the direction of Dr. Roberts. This multisite verification and monitoring approach reduces the need for an independent audit. As such, independent audits will not be conducted.

9.1.8 QUALITY ASSURANCE AND QUALITY CONTROL

Northwestern will serve as the primary site for storage and processing of all study data. However, interventionist and caregiver treatment fidelity data, and child and caregiver caregiverprogress monitoring will be coded by Vanderbilt as needed for clinical support and supervisory feedback to interventionist and caregivers. Linguistic analysis data will be coded and analyzed at the UIUC site. All three sites will enter data using Research Electronic Data Capture (REDCap), a centralized internet-based system, which makes the project database instantly available to all authorized researchers in the system and permits access from any computer with an internet connection from multiple locations (via password protected logon).

Summary reports (e.g., recruitment, assessment progress) can be generated immediately and are automatically updated with additional data. Reports can be created that document participant progress to-date and track missing data. These reports are immediately available to all registered users. At the end of the study, a composite clean data set will be assembled and stored securely at Northwestern and interim data sets will be securely disposed.

Northwestern University will conduct the randomization, oversee data entry procedures, manage the databases, and be responsible for maintaining the quality, integrity, and confidentiality of the database as outlined below. If a breach in the security of the data should occur, the PIs (Roberts, Kaiser, Hadley) will notify the appropriate Institutional Review Board immediately. Northwestern staff will sweep for quality assurance continuously throughout the study under the guidance and supervision of Dr. Siddique. All data will be independently entered by two different research staff. Additional quality assurance checks will be completed to check for missing data, range, and outliers.

Risks related to incorrect scoring of measures will be minimized in the following ways. All standardized and observational measures will be conducted by independent observers and assessors naïve to the experimental condition. All assessments will be administered by trained research staff naïve to the experimental condition, who have been trained to criterion levels on three practice children. All assessments will have a manual with implementation and scoring guidelines. All changes or clarifications made to scoring or administration will be documented in this manual. All testing sessions will be video recorded and procedural fidelity for test administration of observational assessments (i.e., Language Sample, TEGI, Renfrew Bus Story, GFTA) will be measured monthly to assess and ensure accuracy of test administration. Observers for all measures will be trained to point-by-point reliability of .90 prior to coding the observational data. A second independent observer will code 20% of the observational measures to ensure interrater reliability. When coder reliability drops below 80%, a meeting will be held to discuss discrepancies and document additional rules as needed. Staff are trained on data entry procedures using a standard operation procedure. After reviewing the procedures, all staff members must shadow a seasoned research staff member on all data procedures. New research personnel will have all their work checked and verified until their data entry is consistently error-free. Furthermore, scoring of all norm-referenced assessments will be verified by a second person. All reliability and fidelity training and monitoring will be documented in a database that captures each research member's training and monitoring scores. Staff training is also tracked using a project management tool that outlines all training procedures and notifies other research personnel when tasks are completed. If at any time a research member is consistently producing errors, they will be re-trained on all procedures with oversight from the project manager.

Study documents, data, and study procedures will be reviewed by the project manager at set intervals. Every signed consent document will be reviewed by the project manager for proper documentation procedures. All IRB-approved documents will be reviewed for accuracy during a yearly IRB audit. Study data will be reviewed for compliance, completeness, and accuracy using a quality checklist generated by the site principal investigator on a quarterly basis. A report will be generated and shared with the investigators noting any errors or discrepancies with a proposed plan for reconciliation. All errors will be fixed, appropriate documents and standard operation procedures will be amended, and any changes will be shared with the study staff when the discrepancy is identified.

9.1.9 DATA HANDLING AND RECORD KEEPING

9.1.9.1 DATA COLLECTION AND MANAGEMENT RESPONSIBILITIES

Overall cross-site data collection is the responsibility of the staff at Northwestern University under the supervision of Dr. Megan Roberts and Dr. Juned Siddique. They are responsible for ensuring the accuracy, completeness, legibility, and timeliness of the data reported.

Types of study data to be included in this trial include standardized assessments, observational assessments, caregiver report forms, field notes/observations, and video recordings. All data will be stored with their unique identification number. All data will be processed on an ongoing basis as data are captured, within 48 hours of completion. Following upload, data entry, and verification, all the electronic files on the audio and video recorders will be deleted. All study data are stored using: (a) REDCap (Research Electronic Data Capture), a secure, web-based system for collecting and managing data, and (b) a secure password-protected server designated only for study personnel.

Assessment scores for standardized measures will be recorded on a paper copy of the official testing protocol. Research personnel will complete the form at the testing visit or immediately after, and the scores will later be entered into REDCap and uploaded to REDCap and the server. All video recordings will be stored on the server, backed up to an external hard drive, and then deleted from the device. Field notes and observations, as well as protocol deviations and adverse events, will be recorded by the clinician directly into REDCap.

Two specific assessments, known as the language sample and the caregiver-child interaction are study-specific measures. In the caregiver-child interaction, the caregiver and child play with a standard set of toys and books for 30 minutes in their home. In the Language Sample, the assessor engages the caregiver and child in a structured task designed to elicit sentence descriptions of animated events displayed on the iPad. Short written prompts for the caregiver are displayed with each 10-sec animation. In the Language Sample, the assessor obtains the language sample using a standard protocol. These interactions are later transcribed and coded using Systematic Analysis of Language Transcripts (SALT), which automatically generates number of different word roots and total number of words, and the data will be coded using a project-specific coding scheme. Transcripts will be verified by a second transcriber before analysis. This linguistic data coding and analysis will be overseen by Dr. Hadley with assistance from Dr. Rispoli and consultant Dr. Rowe.

Caregiver report forms include several published measures and several study-specific measures that capture basic demographic information and information about the child's additional therapy and services received in the community, as well as child language and behavior. Participant surveys will be completed using an electronic link, which allows the participant to complete their surveys online via REDCap. All survey data is stored by electronic ID number. If the caregiver is unable to access equipment to complete their surveys online they will be offered paper copies, which will later be entered and verified.

9.1.9.2 STUDY RECORDS RETENTION

Study documents will be retained for a minimum of 7 years after the last data are collected on the last research participant. These documents should be retained for a longer period, however, if required by local regulations. No records will be destroyed without the written consent of the sponsor, if applicable. It is the responsibility of the sponsor to inform the investigator when these documents no longer need to be retained.

9.1.10 PROTOCOL DEVIATIONS

All study personnel are trained in the beginning of the study as to what constitutes a protocol deviation. Any procedure that is not outlined in the IRB-approved protocol or in the specific standardized assessment will be considered a protocol deviation. To track and detect these deviations, each clinician or study personnel conducting study procedures is required to complete an ongoing log in REDCap after each assessment and intervention session. This log prompts the study personnel about specific protocol deviations. All research personnel are instructed to inform the principal investigator of any deviations within 24 hours. The project manager at Northwestern University will also create a report in REDCap that will be monitored on a weekly basis to address cross-site protocol deviations, and research members will be updated at weekly staff meeting about how to prevent protocol deviations.

9.1.11 PUBLICATION AND DATA SHARING POLICY

The proposed research will include data from 108 young children with receptive and expressive language delays and their caregivers from two urban sites. The final dataset will include standardized cognitive and language assessments, caregiver-reported demographic and behavioral data, observational data of the children and their caregivers in the clinic and at home over time (19 months). Data from both treatment and control participants will be included. The final dataset will be stripped of identifiers prior to release for sharing; however, we believe that there remains the possibility of deductive disclosure of subjects with unusual characteristics. Thus, we will make the data and associated documentation available to users only under a data-sharing agreement that provides for a commitment: (1) to use the data only for research purposes and not to identify any individual participant; (2) to secure the data using appropriate computer technology; (3) to destroy or return the data after analyses are completed; and (4) to cite the grant and key publications describing the database and measures in any resulting presentations and publications. Transcripts from caregiver-child and examiner-child language samples will also be donated to the Child Language Data Exchange System (<http://chilides.psy.cmu.edu>), an international repository of child language data. Access to the data will be openly available. Researchers who access the transcripts will be expected to abide by the established guidelines for use of TalkBank data <http://talkbank.org/share/irb/options.html>.

Data entry will be completed using a centralized internet-based system (REDCap), which permits multisite access from any computer with a LAN connection from multiple locations with no special software needed beyond internet access (via password protected log on). Northwestern University will manage the database and who has access to it. Vanderbilt University and University of Illinois will have access to the database for data entry purposes. The database will be overseen by Dr. Juned Siddique.

In general, we will follow the guidelines outlined by the American Psychological Association (APA, 2018). Authorship credit will reflect the individual's contribution to the study. An author is considered anyone involved in the following activities: (a) conceptualizing or refining the

research idea, (b) performing a literature search, (c) creating a research design, (d) selecting measures or variables, (e) selecting statistical tests/analyses, (f) performing statistical analyses and computations, (g) interpretation of statistical analyses, (h) writing the manuscript, and (i) responding to reviewer feedback. The primary author will assume responsibility for the publication, making sure that the data are accurate, that all deserving authors have been credited, that all authors have given their approval to the final draft; and handling responses to inquiries after the manuscript is published. Quarterly, the Executive Committee (consisting of the three PIs, Dr. Rispoli, and Dr. Siddique) will identify potential papers and conference submission and respond to any requests to develop a paper or conference presentation. For each paper or presentation, the proposed lead author will complete a preliminary authorship agreement (<http://www.apa.org/science/leadership/students/authorship-agreement.pdf>). We also will use the authorship determination scorecard to determine order of authorship (<http://www.apa.org/science/leadership/students/authorship-determination-scorecard.pdf>) and the APA publication/presentation contract (<http://www.apa.org/science/leadership/students/authorship-agreement.pdf>) to establish a publication timeline for each manuscript. Prior to the start of any publication, the Executive Committee will approve the project. Once approved, the PIs will coordinate any proposed papers or conference presentations originating in their lab, including papers from students, postdocs, and staff. PIs will take responsibility for completing the draft and final author agreements. Preliminary and final author agreements will be approved by Executive Committee, and record of these agreements will be scanned and maintained throughout the project. Only complete baseline data (n=108) may be used in any manuscript prior to the publication of the study aims. Manuscripts presenting the results of all study aims and the analyses listed above must be submitted for publication before any other analyses using the longitudinal data may be submitted.

9.1.12 CONFLICT OF INTEREST POLICY

All research staff members are required to submit a conflict of interest disclosure through their respective universities on an annual basis. Each staff member's disclosure is then reviewed and verified by the staff member's supervisor or the site principal investigator. If a conflict of interest does arise, the site principal investigator, the institutional review board, and the funding agency will be notified to discuss the person's continuation on the project and/or a plan for oversight.

9.2 ADDITIONAL CONSIDERATIONS

N/A

9.3 ABBREVIATIONS

AE	Adverse Event
DLD	Developmental Language Delay
DSMB	Data and Safety Monitoring Board
EMT-SF	Enhanced Milieu Teaching-Sentence Focus
GCP	Good Clinical Practice
IRB	Institutional Review Board
ITT	Intention-To-Treat
NIH	National Institutes of Health

NIH IC	NIH Institute or Center
OHRP	Office for Human Research Protections
PI	Principal Investigator
QA	Quality Assurance
QC	Quality Control
RE	Receptive/Expressive
REDCap	Record Electronic Data Capture
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SMC	Safety Monitoring Committee
SOA	Schedule of Activities
SOP	Standard Operating Procedure
UP	Unanticipated Problem

9.4 PROTOCOL AMENDMENT HISTORY

The table below is intended to capture changes of IRB-approved versions of the protocol, including a description of the change and rationale. A Summary of Changes table for the current amendment is located in the Protocol Title Page.

IRB Modification Number	Approved IRB Modification Date	Clinical Trial Protocol Date	Revisions Made <u>Only</u> to the IRB Protocol	Revisions to Made to the Clinical Trial <u>and</u> IRB Protocol	Affected Section(s) of Changed Clinical Trial Protocol	Rationale for Clinical Trial Protocol Changes
1	06/11/2018	12/9/2019	Added new study personnel	n/a	None	New study personnel
2	n/a	12/9/2019	This modification was discarded. No changes were made.	n/a	None	n/a
3	01/23/2019	01/23/2019	- Added all study personnel for all sites - Added consent forms - Added recruitment materials - Attached copies of surveys - Attached IAA agreements from VU and UIUC	We re-distributed the amount of books and toys from \$75 worth of books and toys at 30, 33, 36, 39, 42, and 45 months to \$150 at 30 months and \$50 at 33, 36, 39, 42, and 45 months.	4.9 (Maintaining Participation in the study)	Caregivers need access to more toys at the start of intervention.
4	02/08/2019	01/23/2019	Added new study personnel.	n/a	None	n/a
5	02/19/2019	01/23/2019	Added new study personnel.	n/a	None	n/a
6	04/30/2019	04/30/2019	Added a QR code to study flyers	Added a table clarifying payment per timepoint	4.9 (Maintaining Participation in the study)	Provide additional detail about payment at each study visit
			Added mosio texting platform for participant scheduling	Added the ADOS as a screening measure	7.2 (Safety and Other Assessments)	Children who fail the STAT screener or there are caregiver and clinician concerns of autism will receive the ADOS. The ADOS is a better measure of autism, but more time intensive to administer. Thus, we will only administer the ADOS in these instances as a safeguard against excluding children who don't have autism.
7	04/03/2019	04/30/2019	Added new study personnel.	n/a	None	New study staff
8	05/15/2019	04/30/2019	- Added "research match" for recruitment - Added that developmental reports will be sent to caregivers by email	n/a	None	n/a
9	06/07/2019	06/07/2019	Changed caregiver materials and consent to reflect the final time point as 49 months rather than 48 months due to two assessment visits at the final time point.	n/a	None	n/a
10	06/05/2019	04/30/2019	Added new study personnel.	n/a	None	n/a

IRB Modification Number	Approved IRB Modification Date	Clinical Trial Protocol Date	Revisions Made <u>Only</u> to the IRB Protocol	Revisions to Made to the Clinical Trial <u>and</u> IRB Protocol	Affected Section(s) of Changed Clinical Trial Protocol	Rationale for Clinical Trial Protocol Changes
11	06/08/2019	06/07/2019	Corrected an error with added study personnel (two people with same exact name)	n/a	None	n/a
12	06/12/2019	06/07/2019	Added new study personnel	n/a	None	n/a
13	06/23/2019	06/23/2019	Updated consent form to get ADOS scores from NU clinic	Added MCDI to screening process rather than at baseline and at T33 and T36 months	1.3 (Schedule of Activities)	MCDI is now used as part of new inclusion criteria and thus required at time of screening.
				Changed inclusion criteria to: (a) PLS AC score of 95 or less AND (b) PLS EC score of 80 or less OR below 10 th % on 2 of 3 sections of MCDI	4.5 (Inclusion Criteria)	The PLS scores were under-identifying children who are at significant risk for DLD. *Approved by the DSMB on 6/21/2019
				Changed intervention terms to match terms used with caregivers and in intervention SOPs	5.12 (Dosing and Administration)	This change ensures consistency across study documents.
				Clarified age of entry is 30-31 months	4.5 (Inclusion Criteria)	It was unclear the exact age at which a child was no longer eligible for the study.
14	06/26/2019	06/23/2019	Small typo in Vanderbilt consent form	n/a	None	n/a
15	07/12/2019	06/23/2019	Added new study personnel.	n/a	None	n/a
16	07/23/2019	07/22/2019	Added language at the beginning of the IRB protocol to denote the clinical trial version date to facilitate cross protocol (IRB, Clinical Trial) consistency.	n/a	None	Obtain official IRB approval of the 7/22/2018 clinical trial protocol.
17	7/31/19	7/22/2019	Added new study personnel	n/a	None	New study personnel
18	8/22/19	7/22/19	Added optional element to NU consent to share PLS-5 at T30 data if participant is also participating in STU208022	Added optional element to NU consent to share data if participant is also participating in STU208022	Section 9.1.1.1 (9.1.1.1 Consent/asses and other informational documents provided to participants)	This ensures that participants who want to participate in both studies (STU208022 and this study) are maximizing time within the lab by not repeating the PLS-5 twice.
19	9/26/19	9/18/19	n/a	Removing premature birth as exclusion criteria	Section 4.6 (Exclusion Criteria)	Increase enrollment
20	10/10/19	10/03/19	n/a	Offering caregivers copies of video clips from study sessions on password protected USB drive	Section 4.9. Strategies for Recruitment and Retention	Increase study retention
21	10/10/19	10/03/19	Added new study personnel	n/a	None	n/a
22	10/11/19	10/03/19	Added new study personnel	n/a	None	n/a

IRB Modification Number	Approved IRB Modification Date	Clinical Trial Protocol Date	Revisions Made <u>Only</u> to the IRB Protocol	Revisions to Made to the Clinical Trial <u>and</u> IRB Protocol	Affected Section(s) of Changed Clinical Trial Protocol	Rationale for Clinical Trial Protocol Changes
23	11/5/2019	10/03/19	Added clarification that the MCDI will also be shared with STU208022 if study time points overlap.	n/a	n/a	n/a
24	11/4/2019	10/03/19	Added new study personnel	n/a	None	n/a
25	12/18/19	10/03/19	Changed wording on NU recruitment flyer	n/a	n/a	n/a
26	2/11/2020	10/03/19	Added new study personnel	n/a	None	n/a
27	3/3/2020	02/12/20	Added new Northwestern Study Flyer Noted additional compensation for additional assessment visits	Noted additional compensation for additional assessment visits	Section 4.9. Strategies for Recruitment and Retention	Increasing study retention by compensating families for additional visits, which aligns with current payment system
28 & 29	7/5/2020 and 7/21/2020	06/18/20	Changed fliers based on new study information for telehealth. Changed informed consent to reflect telehealth changes.	Changes protocols to include all changes for tele-health.	Section 1, 4, 5, 7 and 9.	Due to COVID-19, researchers wanted to reduce in-person visits and keep the trial running, so the study has switched to a primary telehealth/virtual model. Updating phases of intervention to match parent materials
30	9/1/2020	8/20/20	Added additional questions to the screening survey to help identify streamline services for children with autism. Removing the NU HIPAA consent.	Added ADI-R at screening if the clinician has concerns about ASD after the child passed the TELE-ASD-PEDS.	Section 1.3, 4.8, and 7.2.	Added
31	9/3/20	8/20/20	Added new study personnel	n/a	None	n/a
32	12/3/20	11/13/20	Updated study flyer and recruitment materials, updated to one combined study consent form, added recruitment video, updated consent to contact to include collecting MCDI data as a part of screening.	Updated protocol to recruit nationally. T49 visit is now virtual	Section 1.3 and 4.9	The virtual T48 assessment is to accommodate families who do not want in-person visits due to COIVD!9, and the national recruitment is to increase sample size
33	1/26/21	11/13/20	Updates made to the study flyer, recruitment survey/screening consent, and recruitment email script	n/a	n/a	n/a
34	2/16/21	11/13/20	Updated national recruitment flyer and consent typo and update email scrips	n/a	n/a	n/a
35	6/23/21	11/13/20	Added new study personnel	n/a	None	n/a
36	8/9/21	11/13/20	Added new study personnel	n/a	None	n/a
37	8/13/21	8/19/21	Fixed consent form typos and added telehealth questions to Parent Efficacy and Parent Knowledge of Language survey	n/a	n/a	Adding questions about intervention satisfaction related to telehealth
38	9/24/21	11/13/20	Added new study personnel	n/a	None	n/a
39	11/4/21	11/13/20	Added new study personnel	n/a	None	n/a
40	6/13/22	11/13/20	Added new study personnel	n/a	Non	n/a

IRB Modification Number	Approved IRB Modification Date	Clinical Trial Protocol Date	Revisions Made <u>Only</u> to the IRB Protocol	Revisions to Made to the Clinical Trial <u>and</u> IRB Protocol	Affected Section(s) of Changed Clinical Trial Protocol	Rationale for Clinical Trial Protocol Changes
41	6/22/22	11/13/20	Added new study personnel	n/a	n/a	n/a
42		11/13/20	Updated protocol and consent to note that exact assessment scores will be included in reports. Updated recruitment survey to include child race.	n/a	n/a	n/a
43	10/12/22	11/13/20	Added/removed study personnel	n/a	n/a	n/a
44	10/18/22	11/13/20	Added/removed study personnel	n/a	n/a	n/a
45	2/13/23	11/13/20	Added/removed study personnel	n/a	n/a	n/a

REFERENCES

- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- Barton, E. E., & Fettig, A. (2013). Parent-implemented interventions for young children with disabilities: A review of fidelity features. *Journal of Early Intervention, 35*(2), 194-219.
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh T., & Consortium C-2. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development. *Terminology*. doi:10.1111/jcpp.12721
- Bowers, J. (1993). The syntax of predication. *Linguistic Inquiry, 24*(4), 591-656.
- Cleave, P. L., Becker, S. D., Curran, M. K., Van Horne, A. J. O., & Fey, M. E. (2015). The efficacy of recasts in language intervention: A systematic review and meta-analysis. *American Journal of Speech-Language Pathology, 24*(2), 237-255.
- Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders—a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry, 46*(2), 128-149.
- Conti-Ramsden, G., Clair, M. C. S., Pickles, A., & Durkin, K. (2012). Developmental trajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: From childhood to adolescence. *Journal of Speech, Language, and Hearing Research, 55*(6), 1716-1735. doi:10.1044/1092-4388(2012/10-0182)
- Conti-Ramsden, G., & Jones, M. (1997). Verb use in specific language impairment. *Journal of Speech, Language, and Hearing Research, 40*(6), 1298-1313.
- Cunningham, J., Quinn, E., Kaiser, A., & Roberts, M. (2016, November). *Exploring the impact of early intervention in toddlerhood on language & literacy skills in Kindergarten*. Poster presented at the annual meeting of the American Speech and Hearing Association, Philadelphia, PA.
- Curtis, P.R., Kaiser, A. P., Estabrook, R., & Roberts, M. Y. (In press). The Longitudinal Effects of Early Language Intervention on Children's Problem Behaviors. *Child Development*. doi:10.1111/cdev.12942
- Dale, P. S., Rice, M. L., Rimfeld, K., & Hayiou-Thomas, M. E. (2018). Grammar clinical marker yields substantial heritability for language impairments in 16-year-old twins. *Journal of Speech, Language, and Hearing Research, 61*(1), 66-78.
- Deboeck, P. R., & Preacher, K. J. (2016). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling, 23*, 61-75.

- Demir, Ö. E., Rowe, M. L., Heller, G., Goldin-Meadow, S., & Levine S. C. (2015). Vocabulary, syntax, and narrative development in typically developing children and children with early unilateral brain injury: Early parental talk about the “there-and-then” matters. *Developmental psychology*, 51(2), 161.
- Dollaghan, C. (2013). Late talker as a clinical category: A critical evaluation. In Rescorla, L. & Dale, P. (Eds.), *Late Talkers: Language Development, Interventions and Outcomes*. Baltimore: Brookes Publishing Co.
- Dunst, C., Williams, A., Trivette, C., Simkus, A., & Hamby, D. (2012). Relationships between inferential book reading strategies and young children's language and literacy competence. *CELLreview*, 5(10), 1-10.
- Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *Journal fo Speech, Language, and Hearing Research*, 60(10), 2935-2948. doi:10.1044/2017_JSLHR-L-16-0310.
- Fletcher, P., & Peters, J. (1984). Characterizing language impairment in children: An exploratory study. *Language Testing*, 1(1), 33-49.
- Glueck, D. H., & Muller, K. E. (2003). Adjusting power for a baseline covariate in linear models. *Statistics in Medicine*, 22(16), 2535-2551. doi.org/10.1002/sim.1341
- Hadley, P. A., Kaiser, A. P., & Roberts, M. Y. (2018, June). *Early intervention and language delay profile influence toddlers' grammatical growth*. Poster presented at the Symposium for Research in Child Language Disorders, Madison, WI.
- Hadley, P., Rispoli, M., & Holt, J. (2017a). Input subject diversity accelerates the growth of tense and agreement: Indirect benefits of a parent-implemented intervention. *Journal of Speech, Language, and Hearing Research*, 60, 2619-2635.
- Hadley, P., Rispoli, M., Holt, J., Papastratakis, T., Hsu, N., Kubalanza, M., & McKenna, M. (2017b). Input subject diversity enhances early grammatical growth: Evidence from a parent-implemented intervention. *Language Learning and Development*, 13, 54-79.
- Hadley, P., McKenna, M., & Rispoli, M. (2018). Sentence diversity in early language development: Recommendations for target selection and progress monitoring. *American Journal of Speech-Language Pathology*, 27, 553-565.
- Hampton, L., Kaiser, A. P., & Roberts, M. Y. (2017). One year outcomes and predictors of language growth for toddlers with language delays: An RCT follow up. *Pediatrics*, 140(5), e20163646
- Hampton, L. H., & Kaiser, A. P. (2016). Intervention effects on spoken-language outcomes for children with autism: A systematic review and meta-analysis. *Journal of Intellectual Disability Research*, 60(5), 444-463.
- Hauck, W. W., Anderson, S., & Marcus, S. M. (1998). Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Control Clinical Trials*, 19, 249-256.

- Heidlage, J., Cunningham, J., Kaiser, A. P., Trivette, C., Barton, E., Frey, J., & Roberts, M. The effects of parent-implemented language interventions on child linguistic outcomes: A meta analysis. *Early Childhood Research Quarterly*, submitted.
- Hernandez, A. V., Steyerberg, E. W., & Habbema, J. D. (2004). Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *Journal of Clinical Epidemiology*, 57(5), 454-460. doi:10.1016/j.clinepi.2003.09.014
- Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American Journal of Speech Language Pathology*, 19(1), 51-65. doi:10.1044/1058-0360(2009/08-0083)
- Law, J., Charlton, J., & Asmussen, K. (2017). *Language as a Child Wellbeing Indicator*. London: Early Intervention Foundation.
- Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 47(4), 924-943. doi:10.1044/1092-4388(2004/069)
- Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a recent systematic review of the literature. *International Journal of Language and Communication Disorders*, 35, 165–188.
- Leech, K., Wei, R., Harring, J. R., & Rowe, M. L. (2018). A brief parent-focused intervention to improve preschoolers' conversational skills and school readiness. *Developmental Psychology*, 54(1), 15-28. doi:10.1037/dev0000411
- Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.
- Levickis, P., Reilly, S., Girolametto, L., Ukoumunne, O. C., & Wake, M. (2014). Maternal Behaviors Promoting Language Acquisition in Slow-to-Talk Toddlers: Prospective Community-Based Study. *Journal of Developmental and Behavioral Pediatrics*, 35(4), 274–281.
- Loeb, D. F., Pye, C., Redmond, S., & Richardson, L. Z. (1996). Eliciting verbs from children with specific language impairment. *American Journal of Speech Language Pathology*, 5(4), 17-30.
- McGregor, K. K., Rohlfing, K. J., Bean, A., Marschner, E. (2009). Gesture as a support for word learning: The case of under. *Journal of Child Language*, 36(4), 807-828. doi:10.1017/S0305000908009173
- McKean, C., Wraith, D., Eadie, P., Cook, F., Mensah, F., & Reilly, S. (2017). Subgroups in language trajectories from 4 to 11 years: the nature and predictors of stable, improving and decreasing language trajectory groups. *Journal of Child Psychology & Psychiatry*, 58, 1081-1091. doi:10.1111/jcpp.12790

- Mol, S. E., Bus, A. G., De Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent-child book readings: A meta-analysis. *Early Education and Development*, 19(1), 7-26. doi:10.1080/10409280701838603
- Norbury, C. F., & Sonuga-Barke, E. (2017), Editorial: New frontiers in the scientific study of developmental language disorders. *Journal of Child Psychology & Psychiatry*, 58, 1065-1067. doi:10.1111/jcpp.12821
- Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech, Language, and Hearing Research*, 38(2), 434-445. doi:10.1044/jshr.3802.434
- Olswang, L. B., Long, S. H., & Fletcher, P. (1997). Verbs in the emergence of word combinations in young children with specific expressive language impairment. *International Journal of Language and Communication Disorders*, 35(2s),15-33. doi:10.1111/j.1460-6984.1997.tb01622.x
- Özçalışkan, Ş., & Goldin-Meadow, S. (2009). When gesture-speech combinations do and do not index linguistic change. *Language and Cognitive Processes*, 24(2), 190. doi:10.1080/01690960801956911
- Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech Language Pathology*, 5(2), 5-21. doi:10.1044/1058-0360.0502.05
- Paul, R. (2000). Predicting outcomes of early expressive language delay: Ethical implications. In Bishop, D. V. M & Leonard, L. B. (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome*. Hove, U.K.: Psychology Press.
- Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, 17(4), 409-441.
- Ramchand, G. C. (2008). *Verb Meaning and the Lexicon: A First Phase Syntax*. Vol 116. Cambridge, MA: Cambridge University Press.
- Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., Conway, L., Eadie, P., Bretherton, L. (2010). Predicting language outcomes at 4 years of age: Findings from early language in Victoria study. *Pediatrics*, 12(6), 1530-1537. doi:10.1542/peds.2010-0254
- Rescorla, L., Mirak, J., & Singh, L. (2000). Vocabulary growth in late talkers: Lexical development from 2;0 to 3;0. *Journal of Child Language*, 27(2), 293-311. doi:10.1017/S030500090000413x
- Rescorla, L., Roberts, J., & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech, Language, and Hearing Research*, 40(3), 556-566. doi:10.1044/jshr.4003.556
- Rice, M. L. (2013). Language growth and genetics of specific language impairment. *International Journal of Speech Language Pathology*, 15(3), 223-233. doi:10.3109/17549507.2013.783113

- Rice, M. L., & Hoffman, L. (2014). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research*, 58(2), 345-359. doi:10.1044/2015_JSLR-L-14-0150
- Rice, M. L., Hoffman, L., & Wexler, K. (2009). Judgments of omitted BE and DO in questions as extended finiteness clinical markers of specific language impairment (SLI) to 15 years: a study of growth and asymptote. *Journal of Speech, Language, and Hearing Research*, 52(6), 1417-33. doi:10.1044/1092-4388(2009/08-0171)
- Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language, and Hearing Research*, 51(2), 394-407. doi:10.1044/1092-4388(2008/029)
- Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research*, 47(4), 816-834. doi:10.1044/1092-4388(2004/061)
- Rice, M. L., Buhr, J. C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders*, 55(1), 33-42. doi:10.1044/jshd.5501.33
- Roberts, M. Y., & Kaiser, A. P. (2015). Early Intervention for Toddlers With Language Delays: A Randomized Controlled Trial. *Pediatrics*, 134(4), 686-693. doi:10.1542/peds.2014-2134
- Roberts, M. Y., Kaiser, A. P., Wolfe, C. E., Bryant, J. D., & Spidalieri, A. M. (2014). Effects of the teach-model-coach-review instructional approach on caregiver use of language support strategies and children's expressive language skills. *Journal of Speech, Language, and Hearing Research*, 57(5), 1851-1869. doi:10.1044/2014_JSLHR-L-13-0113
- Roberts, M. Y., & Kaiser, A. P. (2012). Assessing the effects of a parent-implemented language intervention for children with language impairments using empirical benchmarks: A pilot study. *Journal of Speech, Language, and Hearing Research*, 55(6), 1655-1670. doi:10.1044/1092-4388(2012/11-0236)
- Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech Language Pathology*, 20(3), 180-199. doi:10.1044/1058-0360(2011/10-0055)
- Rowe, M. L. (2012). A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. *Child Development*, 83(5), 1762-1774. doi:10.1111/j.1467-8624.2012.01805.x
- Rudolph, J. M. (2017). Case history risk factors for specific language impairment: A systematic review and meta-analysis. *American Journal of Speech Language Pathology*, 26(3), 991-1010. doi:10.1044/2016_AJSLP-15-0181

- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children committee on the prevention of reading difficulties in young children*. Washington, DC: National Research Council.
- Tomblin, J. B., Zhang, X., Buckwalter, P., O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research*, 46(6), 1283-1296. doi:10.1044/1092-4388(2003/100)
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of Specific Language Impairment in Kindergarten Children. *Journal of Speech, Language, and Hearing Research : Journal of Speech, Language, and Hearing Research*, 40(6), 1245–1260.
- VanderWeele, T. J., & Vansteelandt, S. (2013). Mediation analysis with multiple mediators. *Epidemiol Methods*, 2(1), 95-115. doi:10.1515/em-2012-0010
- Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of School Outcomes Based on Early Language Production and Socioeconomic Factors. *Child Development*, 65, 606-621. doi:10.1111/j.1467-8624.1994.tb00771.x
- Watkins, R. V., Rice, M. L., & Moltz, C. C. (1993). Verb use by language-impaired and normally developing children. *First Language*, 13(37), 133-143. doi:10.1177/014272379301303708
- Whitehouse, A. J., Line, E. A., Watt, H. J., & Bishop, D. V. (2009). Qualitative aspects of developmental language impairment relate to language and literacy outcome in adulthood. *International Journal of Language & Communication Disorders*, 44, 489-510. doi:10.1080/13682820802708080
- Whitehouse, A. J., Watt, H. J., Line, E. A., & Bishop, D. V. (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International Journal of Language & Communication Disorders*, 44, 511-528. doi:10.1080/13682820802708098
- Windfuhr, K. L., Faragher, B., & Conti-Ramsden, G. (2002). Lexical learning skills in young children with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, 37, 415-432. doi:10.1080/1368282021000007758
- Yew, S. G., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry*, 54, 516-524. doi:10.1111/jcpp.12009