NCT ID Number	NCT05099328
Official Title:	Recasting and Book Reading Under Ideal (Dose-controlled) and Typical (Dose-variable) Conditions: The Role of Fidelity and Adherence in Production and Comprehension Outcomes for Children With DLD
Secondary IDs:	R01DC018276 [U.S. NIH Grant/Contract Award Number]
Statistical Plan Last Updated	July 15, 2025

Risks to Human Subjects

****Updated June 28, 2021 – the primary changes are related to the shift to teletherapy and associated changes in hearing screening protocols and risks to loss of confidentiality.**

****Updated to add changes in inclusionary criteria**

****Updated July 15, 2025 – Includes statistical analysis plan.**

Human Subjects Involvement, Characteristics, and Design

The study is a Phase I clinical trial that employs a randomized parallel design. We plan to enroll 160 children diagnosed with Developmental Language Disorder and their caregivers who will provide services to approximately half of these children. We predict that 15% of the children (N=20) will be lost to follow-up, thus we enroll 160, in order to have 140 children complete the entire study and meet the planned enrollment goals. Children will be randomly assigned to one of four treatment conditions, thus we expect 35 participants per condition. Research will be carried out at the University of Delaware and the University of Maryland College Park.

<u>Children with Developmental Language Disorder (N = 160)</u>

Inclusionary & Exclusionary Criteria:

- Age: 4-7 years old
- Primarily English speaking: Less than 20% exposure to another language, per parent report
- Document poor syntax skills: standard score below 85 on the DELV-NR composite, a dialect neutral assessment (Seymour, Roeper, de Villiers, & De Villiers, 2005).
- Nonverbal IQ within typical range: t-score at or above 35 on the DAS, matrices subscale, (Elliott, 2007).
- Hearing within the typical range: Pass screening using SoundScouts (Gumbie et al., 2018; Mealings et al., 2020) or provide reports of passing a hearing screening at school or in a pediatrician's office. *Due to COVID and limitations on in person activities, treatment may proceed even if these reports are not obtained if no one has concerns about hearing and we will track/follow up with families to obtain a clear hearing screening within 12 months of completion of intervention. Documentation of how many children did not complete hearing screenings will be retained and inform interpretation of data.
- No diagnosis of Autism: Cutoff score of 15 on the SCQ, (Rutter, Bailey, & Lord, 2003)
- No diagnosis of significant sensory-motor concerns or significant psychiatric disorders per parent report
- Able to benefit from treatment:
    - Score below 60% correct on experimenter developed elicited production probes of passives and object relative clauses
    - Producing simple transitive sentences (SVO) and MLU of 2.5 on 100 utterance language sample

<u>Caregivers of Children with DLD (N=160)</u>

A single caregiver will be identified at the time of enrollment to receive training in study procedures. In general, the caregiver's race will align with the child's race. Overwhelmingly women are the caregiver who coordinate this type of care, bring children to appointments, and engage in parent provided treatment. That said, we do not restrict this role to parents or to women. We will take care with our materials to encourage the participation of any primary caregiver.

Inclusionary & Exclusionary Criteria:

- Basic literacy skills per self-report
- Willing to participate in caregiver training and caregiver-based treatments if child is assigned to that condition
- Willing to participate in questionnaires and structured interviews during post-test

Study Procedures, Materials, and Potential Risks

Materials.
Screening and pre- and post-test materials will consist of commercially available standardized tests (see inclusionary and exclusionary criteria above) and experimenter developed probes. The experimenter developed probes will utilize illustrations that are child friendly to support the elicitation and comprehension of passives and object relative clauses.

Screening
After providing informed consent, parents will complete a case history form, the FOCUS questionnaire, the SCQ, and the BRIEF. An SLP or research associate will administer standardized and non-standardized assessments, including the DAS-2, matrices subscale, the DELV, the PPVT-5, the EVT-3, and a language sample. All information obtained will be coded with an alphanumeric code to protect confidentiality and stored on secure university servers. A 10-item elicited production probe will be completed at this point to determine if the child may advance to the pre-test stage.

Pre- & Post-Test
Children who qualify as DLD will be administered 2 20-item production probes to elicit passives and relative clauses. They will participate in a hearing screening, They will complete 2 blocks of 24-item comprehension probes (48 items total) to assess online comprehension of active/passive and object/subject relative clauses using a 2AFC eye tracking task administered via PCIbex. Additional language sample data will be collected to ensure sufficiently large samples for analysis. Audio files, transcription, visual fixation data, and scoring will be coded with an alphanumeric code to protect confidentiality and stored on secure university servers.

Treatment
Children will be randomly assigned to one of four conditions: recasts delivered by parents, recasts delivered by the clinician, syntax stories read by the parent, and syntax stories read by the clinician. In the recast condition, adults will interact with children and recast (restate, with a correction) a child's utterance as either a passive or object relative clause, with the goal of recasting 960 times over the course of the study (1 recast/minute, for 60 minutes, 16 sessions). In the syntax story condition, adults will read illustrated stories containing 30 exemplars of the target structure to children at least 32 times for a total of 960 exposures). These treatments will either be carried out by lab staff or by caregivers. During the treatment period, adherence, fidelity and language environment data will be collected using durable, long life audio recorders (LENAs) in conjunction with time logging/attendance records. In the home condition, training will be provided to the caregivers as to how to complete the treatment protocol via zoom. Caregivers will also be asked to log when treatment is carried out to focus the LENA analyses.

Assessment of the constructs within the Theoretical Domains Framework (TDF)
Motivation, self-efficacy, effort, and treatment efficacy will be assessed using rating scales administered to children (e.g., *how fun was today's session? how hard did you work today?*) clinicians (e.g., *how effective did you feel? how many times did you have to provide redirection?*) and caregivers (e.g., *does your child like coming to treatment? how effective did you feel today?*). These will be tailored to align with the treatment and delivery process. In addition, we will assess parent and clinician knowledge of the treatment through written tests of the grammatical targets and techniques used and observations of them providing treatment in the lab. Caregivers will complete a questionnaire designed to assess barriers within the TDF adapted from Huijg et al. (2014), see appendix. Finally, we will conduct structured interviews with parents during the post-test phase using an ethnographic interview style.

Risks
The primary risks associated with the study are loss of privacy or confidentiality, anxiety, and boredom and fatigue. The study procedures are entirely behavioral in nature and the risks do not exceed those experienced in everyday life. There is also the risk of being assigned to the less effective arms of the RCT.

Alternative Treatments
Children may obtain therapy for their language impairment via free publicly available services or via private pay for treatment. We do not discourage this and will actively connect children not receiving services with Child Find and counsel parents about what treatment they can obtain via private pay. Some families may choose to only participate in our study and others may participate in other therapies. We do not restrict this

choice and document it only. It is unlikely that children will be receiving treatment for the particular structures being studied, but where possible, we will document IEP goals that intersect with the targeted structures and take this into account when interpreting the data.

Adequacy of Protection Against Risks

a. Informed Consent and Assent

Participants will reach out to our labs after receiving information about the study from a flier in the community, their speech pathologist, school, or other community broker. Once they contact a lab, we will describe the study to them, proceed through a brief phone screening, and invite participants to schedule an appointment over zoom at a time convenient to the family. We will request that the child, as well as the participating caregiver, and a guardian who can sign consent forms attend (these may or may not be the same people) the first appointment. At the first appointment, we will present information about the timeline and participation required for the study, describe study procedures, and answer any questions about the study. Families will be invited to sign consent forms for both adults and children. Written consent will be obtained from adults on their own behalf and from a parent or guardian of the child on behalf of the child. Informed consent documents will include appropriate information to allow participants to opt in to data sharing as described in the Data/Resource Sharing and Dissemination Plans.

Children under 7 are generally assumed not to be able to give consent or to understand written assent documents. We will obtain verbal and nonverbal assent from the child instead. Activities will be child friendly and we will move at the child's pace, interspersing demanding activities with breaks. We will invite children to participate in activities (*would you like to play a pointing game?*) up to three times. If children decline to participate or indicate via body language that they are not providing assent, we will represent activities at a later time and work with parents to identify appropriate times of day (e.g., after lunch rather than before) and ways of delivering the activities to make them acceptable to the child (on the floor rather than at a table, with appropriate movement breaks, with a caregiver present). Our lab reserves the right to discontinue participation when we judge there to be a lack of assent, even if parents would like them to continue to participate.

b. Protections Against Risk

The study procedures are entirely behavioral in nature and the risks do not exceed those experienced in everyday life.

As with any study, there is a risk of loss of privacy and confidentiality. This is perhaps increased due to the use of the LENA recorders and online data collection methods. We can mitigate against this by training parents to control the LENA device, allowing them to start and pause the recorder as needed. We also use encrypted (HIPAA compliant) versions of zoom to ensure patient privacy is maintained. We disclose to families during the consent process that we are mandatory reporters for child abuse and that we may come to know things that require reporting via the LENA recordings. Confidentiality and privacy are managed within each lab in general using the following strategies. Screenings and training in the treatment provision are provided at locations and times selected by the families. We use alphanumeric codes to identify participants and code all data. Data are stored on university servers that are accessible via using university IDs and two-factor authentication. Servers are encrypted.

Parents may experience some anxiety about their child's performance on assessments for enrollment in the study or about the gains their child makes during the study. Generally, families are referred to our lab and already receiving services, but sometimes a child's language impairment is first diagnosed by us. We provide counseling to the parents and interpret standardized test results. We make referrals to public services as are appropriate. Families may receive a report describing their child's test results that they may share with professionals if they desire. Parents may also experience anxiety about their own performance during the study. We reassure parents about how change is measured during treatment and that learning to do a new thing can be challenging for anyone.

Children may experience boredom and fatigue. We develop child-friendly activities and move at the child's pace.

There is a risk, as in any RCT, that a child will be assigned to the less effective treatment. Freedman (1987) argues that this risk is acceptable as long as equipoise (or an authentic belief of potentially equal benefit from each treatment at the outset) is held. That does not exist here, as we explicitly express a belief in the proposal that the caregiver delivered recast therapy will be less effective. Johnson, Lilford, and Brazier (1991) argue that it is not the individual investigator who must have equipoise but rather the collective clinical community as a whole that must have some uncertainty as to which of multiple treatments is the most effective. In our own (nonsystematic) conversations with practicing clinicians, they firmly express the belief that recasting by anyone will lead to better outcomes than syntax stories. We believe we have made a convincing argument in the other direction. Thus we believe the field is in a state of collective equipoise on this issue and the RCT is ethically justified.

Nonetheless, we acknowledge the risk of being assigned to the least effective treatment is experienced at the level of the individual participant. We mitigate against this risk in the following ways:

1) Individuals are encouraged to continue to access the public services to which they are entitled.
2) Our labs actively assist families in making the public and private contacts and provides written diagnostic reports to assist in family self-advocacy.
3) While the resources do not exist within this trial to give everyone in-lab recast therapy, the resources do exist to provide parents with the at-home materials and training that they did not receive. Thus, for parents in the recast at home or the two in lab conditions, we will offer the materials for syntax stories. For parents who were in the syntax stories or the two in lab conditions, we will also offer them the training for at home recasting. Parents may opt in to these trainings after post testing as an additional benefit.

c. **Vulnerable Subjects, if relevant to your study**. Children are considered vulnerable participants. Their inclusion is justifiable because the questions are about language learning in children with language impairment. The protocol presents minimal risk and some reasonable expectation of direct benefit, given that the treatments have previously been shown to benefit children with DLD (recast therapy) and typical children (book reading and recast therapy). Thus, work with a vulnerable population is justified.

Potential Benefits of the Proposed Research to Research Participants and Others

Recast Therapy. Recasting is a well-known, evidence based treatment for language impairment that has been documented as providing gains in the area of vocabulary, MLU, morpheme accuracy, and complex syntax, with a good effect size (Hedge's *g* = 0.75-1; Cleave, Becker, Curran, Owen Van Horne, & Fey, 2015). It is often used both alone and as part of a larger therapy package, including Focused Stimulation (Girolametto, Pearce, & Weitzman, 1996), Enhanced Milieu Therapy (Kaiser & Roberts, 2013; Kaiser & Hester, 1994), and Dialogic Reading (Whitehurs, Arnold, Epstein, Angeli, Smith, & Fischel, 1995). Efficacy can be enhanced by treating a single target (i.e., focused stimulation rather than general stimulation; Cleave et al., 2015), ensuring adequate input variability (Meyers-Denman & Plante, 2016), and ensuring that the child is fully attending (Plante et al., 2014). A typical effective dose 600-1000 recasts provided over 10-20 hours of treatment. Parents can be trained to provide this treatment (Fey, Krulik, Frome Loeb, & Procter-Williams, 1999; Proctor-Williams, Fey, & Frome Loeb, 2001). Most prior work has focused on morphology. We have published on the use of recast therapy for complex syntax (Curran & Owen Van Horne, 2019) and our own preliminary data, which was focused on complex syntax instruction shows very large effect sizes, d= 1.84, as compared to children treated on other targets.

Syntax Stories Preliminary evidence for the efficacy of syntax stories largely comes from the typical literature. Vasilyeva et al. (2006) exposed 4-year-olds to 520 examples of passives (e.g., *the girl was kicked by the boy*) through 10 36-sentence stories, each containing 25 exemplars. Hesketh, Serratrice, and Ashworth (2015, 2016) carried out similar studies, exposing children to 300 exemplars of coordinated clauses (*and*), adverbial clauses (*when, because*), embedded clauses (*She said "He is happy")* , or reported speech *(She told us that he is happy)* by way of 10 short stories read aloud to the class. In all three studies, TD preschool and school-age children showed improvement on the trained structure. In our own preliminary

data, 9 children with DLD and 9 TD children, 3-5 years old, had syntax stories read to them by caregivers several time a week, for an average of 583 (range 408-816) exposures to each target. Results indicate increased production probe accuracy for treated forms only, with gains evident for both TD and DLD children (Treated-Control, TD: $p = .01$, $d = 1.02$; DLD: $p = .005$, $d = 1.32$). Although TD children make greater absolute gains, the DLD children have lower pre-test scores and more consistent (less variable) gains and thus show larger effect sizes.

Thus, there is reason to expect that children will acquire new grammatical forms and increase their mean length of utterance over the course of the study. The degree of benefit obtained will likely vary by child and by the degree to which parents faithfully implement the treatment. There is less known about degree of benefit via teletherapy but preliminary evidence from our own lab suggests that it is approximately equally effective for children who participate in online activities consistently. Given that risks are minimal, this is a reasonable trade off. At the conclusion of the study, parents will receive the materials used to train the at-home conditions in book reading and recasting if they so desire.

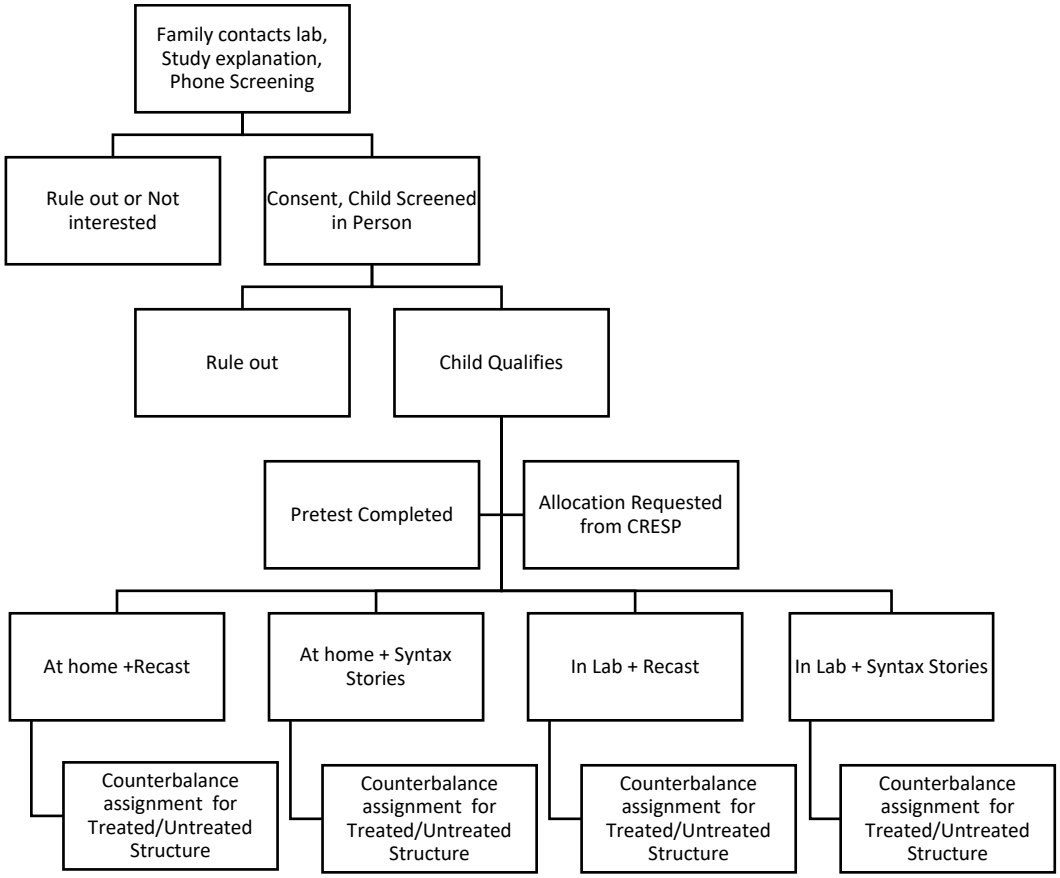
Importance of the Knowledge to be Gained

Understanding how these two different treatments are differently administered by lab staff and parents will inform our understanding of the effectiveness of these treatments under everyday delivery conditions and provide a foundation for improving implementation of treatments to improve grammar outcomes. Linking changes in grammatical processing to treatment outcomes will help us to understand whether commonly used therapies like recasting have an effect on both comprehension and production or only on production. These results have the potential to change how treatment is provided to children affected by DLD (7% of the population).

Statistical Design and Power

Randomization

To guard against bias, children are randomly assigned to condition. A randomization scheme was pre-allocated at the beginning of the study by Henry May at CRESP and uploaded to RedCap using a blinded allocation module. Allocation is requestable as children qualify for the study from CRESP. This ensures that the assessor at pre-test is blind to allocation. To ensure approximately equal allocation to each condition at each site, separate randomization schemes are used for each site. Flow through the study and drop outs at each phase are tracked following EQUATOR reporting guidelines.



Data Reduction

Estimated Dose

We compute estimated dose provided as follows:

*Dose = number of sessions reported by families * percent sessions documented based on recordings * recasts/sentences read verbatim observed in transcribed sessions*

Production Data

The dependent variable is production accuracy scored item by item as correct (1), incorrect (0), or unscorable (missing). Individual items are additionally coded as the syntactic type elicited (passive voice or object relative clause) and as treated/untreated clause type since that is counterbalanced across participants.

Eye-tracking Data

Eye-movements are continuously sampled at the rate of 500 Hz (every 2 ms) from trial onset to the onset of the pointing prompt. Fixations are coded as present (looks to an area of interest on screen) or missing (looks away from these interest areas, track loss, blinking). Present looks are recoded as Agent fixations or Patient fixations based on the looks' relation to target sentences. "Agent fixations" are looks to areas of interest where first-mentioned noun phrases are agents (i.e., correct interpretation for actives and subject relative clauses). "Patient fixations" are looks to areas of interest where first-mentioned noun phrases are patients (i.e., correct interpretation for passives and object relative clauses). Based on pilot data, linguistically mediated eye-movements in responses to the speech stream occur primarily during a 2-second window from the disambiguating verb morphology. Also, consistent with prior work (Huang et al., 2013; Huang et al., 2017), our pilot data indicate that interpretations of canonical structures (i.e., actives) occur faster than noncanonical structures (i.e., passives). On this basis, we will run separate analyses over an "early window" (e.g., first 1000 ms after disambiguation) and "late window" (e.g., second 1000 ms after disambiguation). To examine sensitivity to the timing of window transitions, and to better pinpoint typical timing of stimulus effects, we will explore a range of time values distinguishing potential early and late windows.

Within these windows, we will first calculate **Agent Preference** for each item and then use this value to derive **Structural Sensitivity** for each set of sentences (active/passive, subject/object relative clause). First, for each test point (pre- and post-) and sentence type, we will determine whether canonical and noncanonical structures elicit different interpretations. **Agent Preference** will be calculated as an averaged proportion of Agent looks over Agent plus Patient looks. Values greater than 0.5 indicate a preference to look at the Agent while values less than 0.5 indicate a preference for the Patient.

Second, for each time window (early and late), we will assess the extent to which Treatment Condition affects interpretation of the treated noncanonical structure. Our dependent variable will be **Structural Sensitivity,** calculated as *an averaged difference score of **Agent Preference** in the canonical structure (SRC, active) minus the noncanonical structure for each structure (ORC, passive) for each child*. This measure is selected rather than Agent Preference because it captures the **degree to which a child differentiates the two structures** from each other. Sensitivity to passive/object relative clause cues will be evident in values greater than 0, with higher values mapping onto increased sensitivity. Each child will receive a single Structural Sensitivity score for active/passives and a single Structural Sensitivity score for subject/object relative clauses at each test point.

Analytic Approach

For both comprehension and production outcomes, we utilize generalized linear mixed models to examine change over time for each condition. The adoption of generalized linear mixed models (GLMM) with random effects for subjects and items as the statistical approach has advantages in that we do not need to exclude children for failing to produce scorable responses given that this model can account for uneven response rates across children (Baayen, Davidson, & Bates, 2008) and it accounts for item level variance that can affect results.

Production probe accuracy will be analyzed with a logit link function. Use of a logit link function for production data also closely mirrors the underlying data structure in which individual items are scored as correct or incorrect. Structural sensitivity will be analyzed with linear mixed-effects models. Maximal models will include random intercepts (and possibly random slopes) for subjects, items, and clause type (Barr, Levy, Scheepers, & Tily, 2013), and parameter-specific p-values will be estimated with degrees of freedom determined through Satterthwaite approximation (Luke, 2017).

We will model comprehension and production outcomes separately, though we make similar predictions as to the effect of treatment on each outcome measure. Prior to carrying out the main analyses, we will check for differences attributable to participant sex and race (relevant biological variables). These will be included in future analyses if necessary, but prior work suggests that they are not likely to influence outcomes.

We will also use generalized additive mixed models (GAMMs) to model longitudinal patterns in eyetracking data (i.e., looks to the agent) to examine and visualize timing of stimulus effects and changes in focus at key points during and after presentation of the prompt. GAMMs model fixations as an intensive time series (Cho et al., 2018, 2022) and allow identification of key points of separation. Moreover, the GAMMs models do not require prespecification of windows of time, providing more flexibility and sensitivity when modeling differences in eyetraking over time and between groups. The specification of the GAMMs models will mirror the GLMM equations below, except that smoothing functions will be used for an additional TIME predictor to fit a spline regression for each time series.

Aim 1. Determine the relative differences in the efficacy of illustrated syntax stories and recast therapy when treatment is dose controlled and delivered by highly trained SLPs in the laboratory.

We will first confirm that we achieved control of dose in the two in-lab conditions. We will carry out an unpaired two-tailed t-test comparing the estimated dose delivered for each condition. We anticipate that the two conditions will not differ and that the distributions will show high degrees of overlap ($p>.5$, $d<.3$).

To test the key question in Aim 1, we will compare the relative effect of treatment condition by testing for differences between the treatment effect estimates of stories in the lab and recasts in the lab conditions.

Well-matched groups at pretest will be evident if the two conditions do not differ at pre-test. If one treatment condition is more effective than the other, this should be evidenced by an interaction between test point and treatment contrast effects that are significant at the p<.05 level; if, as we hypothesize, the two conditions are equally effective there will be no significant interaction between test point and treatment. We are powered to detect an effect size difference of .35 (2 raw score points) or larger on production.

*Likelihood of a Correct Production = TestPoint *TreatmentCondition*
+ (1|item)+ (clause type |subject)
*Structural Sensitivity = TestPoint *TreatmentCondition*
+ (clause type |subject)

In the somewhat likely event that the interaction between time point and treatment is not significant, we can directly test for treatment efficacy by comparing the effect of treatment on the treated vs. untreated structures. We expect that the model estimate for Treated or Untreated clause will be significant at p<0.05 level, favoring the treated clauses, demonstrating that treatment was effective and ruling out maturational effects.

*Likelihood of a Correct Production = TestPoint *TreatedOrUntreated*
+ (1|item)+ (clause type |subject)
*Structural Sensitivity = TestPoint *TreatedOrUntreated*
+ (clause type |subject)

Aim 2. Determine whether the influence of delivery method (illustrated syntax stories or recast therapy) on dose delivered (number of exposures to the target) is of sufficient magnitude that changes in child outcomes are observed when treatment is delivered by caregivers.

We will first determine whether dose differs across the two at-home conditions. We will carry out an unpaired two-tailed t-test comparing the estimated dose delivered for each condition. We anticipate that the two conditions will differ and that the distributions will show low degrees of overlap (p<.05, d>.5).

To test the key question in Aim 2, we will test for a dose by test point or dose by treatment condition by test point interaction for stories at home and recasts at home conditions. Well-matched groups at pretest will be evident if the two conditions do not differ at pre-test. If one treatment condition is more effective than the other, this should be evidenced by an interaction between test point and treatment contrast effects that are significant at the p<.05 level; if, as we hypothesize, dose is a critical component of this difference, a two- or three-way interaction between time and dose or between time, treatment and dose will be observed. We would interpret the two-way interaction as indicating that dose, rather than treatment is critical for obtaining the key outcomes. A three- way interaction suggests that the delivery method modifies the effectiveness of the total dose.

*Likelihood of a Correct Production = TestPoint *TreatmentCondition*EstimatedDose*
+ (1|item)+ (clause type |subject)
*Structural Sensitivity = TestPoint *TreatmentCondition*EstimatedDose*
+ (clause type |subject)

In the event that the interaction between time point and treatment is not significant, we can directly test for treatment efficacy by comparing the effect of treatment on the treated or untreated structures. We expect that the model estimate for Treated vs. Untreated clause will be significant at p<0.05 level, favoring the treated clauses, demonstrating that treatment was effective and ruling out maturational effects.

*Likelihood of a Correct Production = TestPoint *TreatedOrUntreated*
+ (1|item)+ (clause type |subject)
*Structural Sensitivity = TestPoint *TreatedOrUntreated*
+ (clause type |subject)

Power Analysis for Production

Power analyses were carried out using Monte Carlo simulation (5,000 iterations) via SAS 9.4 and PROC GLIMMIX using residual pseudolikelihood estimation of a binary logistic mixed effects model with item responses (n=20) nested within tests (n=4 representing test point and clause type repeated measures) that are nested within children (n=120) who have been randomly assigned to treatment conditions, counterbalanced across the 2 clauses. Random effects are included for items within clauses, as well as four correlated child-level random effects (i.e., for each of the four repeated measures). A single model was used to estimate power in alignment with the overall design of the study in which children are randomly assigned to one of FOUR conditions. Power for the more focused models described above can be inferred from the power estimates for the treatment contrasts below. By adopting this randomization scheme and model design, we retain the ability to test unplanned contrasts that go beyond the planned analyses for the Aims if we so chose (e.g., recasts in the lab vs. recasts at home) in order to carry out secondary analysis associated with cost effectiveness, dose, or home environment.

Test scores were simulated based on means and standard deviations of pretest and posttest scores from a previous study. Pretest scores for both clauses were sampled from a Normal distribution with M=3.0 and SD=2.5, rounding to the nearest integer, with a minimum score of 0 and maximum score of 7 items correct (corresponding to the >40% correct study eligibility criteria). Posttest scores for untreated clauses were sampled from a Normal distribution with a correlation of .90 with pretest scores and M=3.0 and SD=2.5, rounding to the nearest integer, with a minimum score of 0. Posttest scores for treated clauses were sampled from a Normal distribution with a correlation of .90 with pretest scores, rounding to the nearest integer, with a minimum score of 0, and M=5.5 and SD=4.0 for the syntax stories at home condition (ES=+1.0), M=4.0 and SD=4.0 for the parent recasts condition (ES=+0.4), M=6.0 and SD=3.5 for the syntax stories in the lab condition (ES=+1.2), and M=8.0 and SD=3.5 for the lab recasts condition (ES=+2.0).

Based on the above effect size estimates, power for detecting main effects in Aim 1 exceeds .95 for the syntax stories at home, syntax stories in the lab, and lab recasts. Power is .61 for detecting main effects of parent recasts (i.e., the effect of parent recasts would need to exceed .50 to be detected with 80% power). Power for treatment effect contrasts is .75 to detect differences in syntax stories at home and parent recasts effects, .90 to detect differences in parent recasts and syntax stories in the lab effects, and .99 to detect differences in parent recasts and lab recasts effects. Other contrasts (e.g., syntax stories at home vs. syntax stories in the lab) have power less than 80% if the expected effect sizes hold true (i.e., +1.0 vs. +1.2); differences in effects would need to exceed 0.35 logits (2 raw score points) in order reach statistical significance with 80% power.

The influence of dose (Aim 2) was simulated by predicting different dose distributions for each condition. These distributions were simulated and then we randomly sampled from the distribution for each case. Hypothesized distributions of dose lead to substantial changes in the result, and thus we make our method of construction these distributions clear here. We assume that a minimum dose is achieved during the training visits (home condition) or by attending at least 1 session (lab condition) for all children. The two in-lab conditions are highly similar and both assume that 25% of participants attend between 1 and 12 sessions, 25% attend between 12 and 16, and the remainder have perfect attendance. This is a conservative estimate based on attendance data from our summer camp study.

The two at-home conditions require slightly more explanation. We assume for recasts that the first quartile has a dose of up to 1 recast/5 min for 60 min per week. The next quartile extends up to 1 recast/5 min for 120 min per week. The third quartile ranges up to a dose of 1 recast every 3 min for 120 min per week and the last quartile assumes target delivery. Similarly, syntax stories at home escalates from reading 1x/week to reading 2x/week to reading 4x/week (as asked) to reading 1 syntax story daily as the maximum. Several families in the pilot data for the syntax stories read the provided syntax stories daily.

Table 1. Hypothesized dose delivered by condition, in quartiles

	Recasts in lab	Stories in lab	Recasts at home	Stories at home
minimum	60	60	60	60
1st quartile	768	768	96	240
median	912	930	192	460
3rd quartile	960	960	320	960
maximum	1008	960	960	1680

Next the gain obtained from each dose was estimated by multiplying the dose provided by a value drawn from a uniform distribution ranging from .0005 to .006, to account for the fact that some children will be more responsive to any given dose than others. Finally, the gain was added to the pretest scores to yield posttest values. Simulations were checked to ensure that final post-test means and standard deviations for the treated structure were in line with estimates for post-test gains used for Aim 1 simulations. Planned statistical analysis were carried out over the simulated data and these simulations were repeated 1000 times.

Simulations that include dose indicate power of .93 for models that include dose and time. Effects of treatment, dose, and power combined are not detectable because this estimation of dose confounds the distribution of dose with treatment delivery approach, as we hypothesize in the strategy. It suggests that three-way interactions will only be detectible if the delivery method modifies the benefit obtained from each exposure to the target form.

Power Analysis for Eye Tracking

Power analyses were carried out using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). Our preliminary data focused on eye-movements during a single test point, thus we do not have relevant estimates of fixation changes at multiple time points or following interventions. Consequently, we based our sample size on power analyses for production (n=120) and calculated the smallest changes in Structural Sensitivity that will be detectable with varying levels of power.

In our preliminary data (n=15), children with DLD revealed M = 0.02 and SD = 0.17 in the late window. With the proposed sample size, effect sizes of d=.30 will be detectible with 80% power, d=.34 with 90% power, and d=.42 with 99% power. These correspond to increases in Structural Sensitivity of M=.05, M = .06, and M = .07 respectively. For reference the TD children in our Preliminary Data have a Structural Sensitivity score that is M=.29 larger than that of the children with DLD. Based on this benchmark, improvements across time points that are 17-24% in magnitude would be detectable in the current design. This suggests that this study is adequately powered to detect relatively small changes in linguistically mediated fixations.