Learning science concepts alongside language goals: A telehealth replication of a randomised controlled trial examining whether children with developmental language disorder can learn vocabulary or grammar in combination with curricular science content

Lindsey Hiebert, Maura Curran, Karla McGregor & Amanda Owen Van Horne

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Abstract

Purpose: We asked whether children with Developmental Language Disorder (DLD) can learn vocabulary or grammar targets and curricular content simultaneously. We replicated prior work integrating two language interventions into a first-grade science curriculum and extended it by testing delivery via teletherapy.

Method: A parallel arm randomised controlled trial was conducted-using telehealth approaches. Four-to-7-year-olds with DLD were randomly assigned to one of three arms, science-only (n=13), science + grammar (n=11), and science + vocabulary (n=10), with fidelity documented for both science and language instruction. The primary outcome measures were changes in the taught language targets and science content, with secondary outcomes including distal measures of language and science.

Result: Complete data for 32 participants were analysed with mixed effects regression. All arms improved on science and grammar targets, with gains in the vocabulary arm exceeding those in the control arm. There were no gains on the distal measures.

Conclusion: Similar to the findings in the replicated study, children with DLD can learn language targets in the context of curricular instruction. Enhanced rich vocabulary instruction holds-promise as an approach that can be embedded in the curriculum and produces gains both in person and via telehealth methods of instruction.

Keywords: DLD, telehealth, grammar, vocabulary, science

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Developmental Language Disorder (DLD) is a neurodevelopmental condition that limits language learning, comprehension, and expression, thus presenting barriers to academic success (Duff et al., 2022). Language therapy is frequently provided in the school setting via a pull-out model in the United States, causing students with DLD to miss general instruction time (Finestack & Satterlund, 2018). Integrating language therapy into the instructional setting in a way that supports both language goals and access to the curricular content seems the ideal solution (Archibald, 2017). Although early efficacy data support the embedding of vocabulary and grammar targets in science curricula (Curran & Owen Van Horne, 2019; McGregor et al., 2021; Owen Van Horne et al., 2023), additional evidence is needed to demonstrate that children can learn both a language target and curricular content simultaneously. In this paper, we sought such evidence. We replicated prior work examining the integration of two language interventions into a first-grade science curriculum and extended this approach by testing delivery via teletherapy.

Rationale for Telehealth Methods

Although tele-platform language therapies have been trialed for rural and other hard-to-reach populations (e.g., Sutherland et al., 2016; 2017), the COVID-19 pandemic greatly accelerated their use (e.g., Campbell & Goldstein, 2022; Peña & Sutherland, 2022). Telehealth approaches to child language assessment (Pratt et al., 2022; Schmitt et al., 2022; Sutherland et al., 2021) and intervention (Arnold et al., 2021; Castilla-Earls & Owen Van Horne, 2023) appear comparable to in-person methods across a variety of populations. Relevant adaptations for tele-platform assessment include ensuring a caregiver is present but instructed against influencing results, and devising ways to share visual materials effectively (Pratt et al., 2022; Schmitt et al., 2022; Sutherland et al., 2021).

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Science and Language Intervention

One can only teach or learn science by engaging in rich communicative exchange, and the natural synergy between the two domains may be exploited for the benefit of students. For example, following coaching focused on improving science instruction, teachers used open-ended questions more frequently than before, and children took more and longer turns, produced longer utterances, and included more complex syntactic forms in their responses (Menninga et al., 2017b). Growth in grammar (Menninga et al., 2017a) and vocabulary skills (Menninga et al., 2017b) was observed during science activities for monolingual children, and comparable vocabulary outcomes have been documented for bilingual children (Restrepo et al., 2013).

In a series of early efficacy studies, we evaluated science lessons as a context for embedding language therapy for children with DLD. Curran and Owen Van Horne (2019) provided proof of concept in a Single Case Experimental Design study that demonstrated learning of *because* clauses but not *so* clauses. In group designs, we found a benefit for learning complement clauses (Owen Van Horne et al., 2023) and academic vocabulary (McGregor et al., 2021) in some but not all students with DLD. The results reported by McGregor et al. (2021) and Owen Van Horne et al. (2023) derive from a single data set collected from children with DLD who attended a summer science and language camp. Children made gains in both the language and science targets from pre- to post-test, with improved complement clause expression being the most robust outcome. The COVID-19 pandemic necessitated the discontinuation of the study before we could gather enough data to compare science outcomes in the language treatment arms to science outcomes to a business-as-usual control arm.

Current Study

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

This study replicates the core active ingredients of McGregor et al. (2021) and Owen Van Horne et al. (2023) while extending the findings to a telehealth platform. Key similarities to the prior work are 1) a focus on children with DLD, 2) the use of a popular first grade physical science curriculum aligned with national standards, the Full Option Science System (FOSS) Next Generation Science Standards (NGSS) (Lawrence Hall of Science, 2015), 3) the inclusion of grammar and vocabulary intervention arms, and 4) intervention provided to children in small groups.

There are fundamental differences as well. The prior study was organised as a 5-week, in-person summer camp with 2-hour daily lessons, for a total of 38 hours of instruction. Pre-service SLPs taught physical and life sciences units with grammar, vocabulary, or phonological awareness interventions added to the science instruction. The current study focused exclusively on physical sciences and was condensed to a 6-week 2-day per week (~6-9 hours) intervention provided by certified SLPs via a telehealth approach, a more scalable amount of intervention than previously offered. Rather than a phonological awareness arm, science-only instruction was used as the control arm. All instruction in the telehealth version was augmented by pre-recorded books. Our core predictions were:

- Children will learn science content in all three arms, and the extent of learning will not differ between arms.
- Children's learning of language targets will be specific to arm: There will gains in the use of the target sentence structure in the grammar arm and gains in the comprehension of target words in the vocabulary arm, but no gains in the control arm,

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

- Children in the language treatment arms will demonstrate generalised gains in language (number of content words and thematic units included) during a science retell task, with greater gains evident when the stories focus on trained science topics.

Method

The study was a randomised controlled trial using a parallel arm design with three arms. Aside from the under-recruitment of participants, there were no deviations from the planned protocol. Sample size was determined from a power analysis conducted in association with the original in person trial (target n= 17 per condition).

Participants

Participants were recruited between December 2020 and January 2023, with post-testing for the last recruited participant in April 2023. Social media ads, word-of-mouth, and flyers were used to reach families who had concerns about their children's language development. Interested families were screened for eligibility to proceed to a consent meeting on the basis of living in the USA, having access to appropriate technology, and having a child with speech/language concerns between the ages of 4 and 7 years old. Parents or legal guardians provided written informed consent for the child to participate prior to the beginning of screening and children provided ongoing verbal assent throughout the protocol. The CONSORT diagram (Figure 1) illustrates the number of children at each stage from recruitment to completion. Thirty-two participants completed the entire treatment protocol: 13 in the science-only arm, 10 in the science + grammar arm, and nine in the science + vocabulary arm.

We recruited children who had not begun first grade to ensure no prior exposure to the first-grade curriculum we used. Other inclusion criteria were: English as a primary language per parent report on a structured questionnaire (MAPLE: A Multilingual Approach to Parent

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Language Estimates, Byers-Heinlein, et al., 2020), a standard score below 85 on the Diagnostic Evaluation of Language Variation-Norm Referenced (DELV-NR; Seymour, Roeper, De Villiers, & De Villiers, 2005), a nonverbal IQ within the typical range as demonstrated by a t-score of 35 or greater on the Differential Abilities Scale-II, matrices subtest (DAS-2, matrices, Elliott et al., 2007) and hearing within the typical range, as documented by SoundScouts (Dillon et al., 2018; Mealings et al., 2020). To ensure that a child had the potential to benefit from the intervention, children were included if they could produce simple 3-5-word sentences with at least 50% intelligibility (based on clinical judgement and language samples), as well as finite complement clauses, a target form, with 40% or lower accuracy on a 10-item elicited production probe. Children with a diagnosis of autism, Down syndrome, traumatic brain injury, or intellectual disability based on parent reports were excluded. Additional details appear in Table 1.

Pre- and Post-Testing

Children in all arms participated in pre- and post-test treatment measures that involved 1) 20 attempts at complement clause productions elicited via video clips, 2) responses to a 20-item 3-alternative-forced-choice task that tapped comprehension of target and non-target words, and 3) a science test for each science unit. We also employed a retell task designed as a distal measure of science and language gains. The retell items were six passages that tapped scientific topics that were taught (air, sound, light) or untaught (legs, dolphins, levers). Following guidelines adapted from Mantzicopoulos and Patrick (2010), we scored each passage for the number of content words and thematic units that the child included in the retelling. None of the content words had been directly targeted during teaching. For example, consider the Air Passage: *"Windy weather is when air moves around quickly. Windy weather can blow things all around. If you have an umbrella during windy weather, it could blow away. A little bit of wind can feel nice*

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

on a hot day." There are 17 unique content words (underlined above) and four thematic units (wind is moving air; wind blows things; an umbrella is one of those things; wind feels nice). If a child retold the story as "wind blows stuff," they would be credited with two content words and one thematic unit. The detailed procedures for each test and the rubric for scoring distal retells appear in the Supplemental Materials.

Participants completed pretesting within an average of 16 days (range 7-42 days) before the treatment and post-testing within an average of 8 days (range 2-18 days) after the end of treatment. The children participated via Zoom from their own homes but were still able to interact with the others in their group online. They attended a total of 89% of all teletherapy sessions. All testing procedures were conducted by licensed SLPs with more than two years of experience. When scheduling allowed, SLPs that had not provided the intervention conducted post-testing (22/35 post-testing masked).

Random Assignment

A core feature of inquiry-based science instruction practices is the need for multiple learners to participate together. Thus, once a child qualified to participate, we actively sought one or two additional qualified children with similar schedule constraints who would be assigned to the same arm as a group. The preference was for groups of three but sometimes, due to family schedules, groups of two were formed in order to move families forward in the research process more quickly. The fourth author, who remained masked to participant demographic characteristics and test scores, randomly allocated groups to arms. Using Sealed Envelope (2020), she created a randomised list of blocks sized three, six, or nine. Within each block, the three treatment arms were equally represented.

Treatment Arms

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Science-Only Control. Three physical science units (air, sounds, and light) adapted from the FOSS NGSS curriculum developed by Lawrence Hall of Science (2015) were taught in two 30-to-45-minute sessions per week for six weeks. The FOSS NGSS curriculum is a popular science curriculum used in approximately one-third of all schools serving grades K-12 across the United States. Following NGSS, it uses an inquiry-based learning approach in which students engage in the practice of science to explore critical phenomena and then engage in sense making activities to come to understand scientific principles. Students are expected to learn about core disciplinary content (e.g., sound is made of vibrations), scientific practices (e.g., to make observations and keep data), and cross-cutting concepts (e.g., cause-effect) through guided participation in various types of activities. Each science unit included four hands-on lessons with materials (e.g., tuning fork, flashlight, test tubes) shipped to families before the treatment. Visual supports for science instruction included a visual schedule, graphics adapted from FOSS materials, and short videos of some phenomena (e.g., a hot air balloon). Science learning was tested via Science I-check questions adapted from the FOSS curriculum. Adaptations to the I-checks were described in Owen Van Horne et al. (2023) and include things like allowing pointing responses rather than labeling and reducing the number of clauses in the questions. Sample lesson plans for each unit appear in the Supplemental Materials.

In consultation with Lawrence Hall of Science advisors, we adapted the curriculum to make it accessible to preschoolers, as described in Owen Van Horne et al. (2023). To support teletherapy, we also augmented the curriculum with video-taped read aloud passages from age-appropriate science books. These books included visuals and descriptions of content that was difficult to complete in the session, such as a discussion of sounds one might hear in nature or how the sun serves as a light source to form shadows outside. The result was 36 read-alouds

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

(video length M = 6.73 min, SD = 1.89), which parents and children watched at a rate of six per week outside of the therapy sessions. The electronic platform logged the time the book began playing and the time that an attention-check question (e.g., what color was the balloon in the book?) was answered. The combination of the accuracy of the attention check question and the duration of play allowed us to gauge compliance. Thus, we were able to be faithful to the scope of content and the connection between natural phenomena and science lessons while still working on a remote teletherapy platform. A list of books by unit appears in the Supplemental Materials.

Grammar. Finite complement clauses (e.g., We thought that it would float) were the target structure in the grammar condition. Growth was supported by various evidenced-based therapy techniques likely to work in concert with each other, including recasting, modeling, elicited production, explicit instruction, and visual supports. In each session, the SLP provided five recasts and three elicited production opportunities for each participant in each session. For example, the clinician might say, *"What do you think will happen?"* A child might respond with, *"It float*." The clinician would then recast with, *"You thought that it would float."* In addition, each small group was exposed to between 35 to 40 models, one elicited imitation *"please repeat the sentence,"* and three elicitation attempts/models using visual supports that highlighted the need for a main clause, a subordinating conjunction (that) and a subordinate clause. The at-home read-alouds for children assigned to the science + grammar arm were rewritten to include 20 complement clauses in each story.

Vocabulary. The vocabulary treatment combined the methods described above for the control arm with enhanced Robust Vocabulary Instruction (Beck et al., 2013). Twelve vocabulary words (*compare, diagram, evidence, explanation, hypothesis, materials, model, multiple, pattern,*

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

problem, scientist, search) were distributed evenly throughout the science lessons (two words per lesson, with each word presented in two different lessons, once in the first three weeks and once in the second three weeks). At a minimum, the SLP modeled each target word in context three times, provided one child-friendly definition to the group, and elicited two production attempts from each participant. She also emphasised the forms of the new words by highlighting their initial sounds, clapping out their syllable structure, and comparing them to similar-sounding familiar words. The emphasis on form constituted an enhancement to the Robust Teaching techniques, which focus on word meaning. At the end of each science lesson, the target words embedded therein were reviewed. For example, the clinician might say, "*Today we learned about air. Air is everywhere! We also learned the word hypothesis – let's each say it!"* The at-home read-alouds for the science + vocabulary arm were altered so that the child would hear each target word twelve additional times in an informative context.

Treatment Fidelity and Reliability

The fidelity of the science teaching and language treatment elements, techniques, and dosages was excellent and, with some exceptions in terms of grammar and vocabulary models (see Table 2), the SLPs avoided cross-contamination of the treatment arms by avoiding vocabulary- or grammar-specific treatment techniques and minimising models of target vocabulary and complement clauses in unassigned arms (Table 2).

To determine the fidelity of at-home procedures, we measured the number of log-ins to the video read-alouds and the accuracy of responses to attention check questions. Low accuracy of attention checks, combined with wide standard deviations around logins suggest that the vocabulary group was not as compliant as the other groups (Science: Videos Started M = 22.4, SD = 7.8; Accuracy on attention check M = 92% SD = 7.4%; Grammar: Videos Started: M = 25,

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

SD = 7.2; Accuracy on attention check: M = 83.8%, SD 13%; Vocabulary: Videos Started M = 17, SD = 12.1; Accuracy on attention check M = 71.3%, SD 37.7%).

Analytic Approach

All analyses were conducted using R package lme4 with syntax derived from Meier (2022) and Zhang and Wang (2017). Following the preregistration, a series of mixed-effects regression models that included time, treatment arm, and interaction between treatment arm and time were tested to determine whether the post-test scores were higher than pre-test scores and whether participation in a specified arm showed greater improvement on that particular test. In addition, a second mixed effect model for each assessment that only tested for pre-test differences across condition was conducted. Individual models were created for each type of assessment (Science I-check, complement clause production probes, vocabulary comprehension probes, and science measures). Tables 4 and 5 depict the science-only control group as the arm with which the grammar and vocabulary arms (conditions) were compared. Fully interpreting these models requires shifting the reference variable to complete pair-wise comparison across conditions. Models using grammar and vocabulary as reference variables are available in the supplemental materials. Logistic models were used for Science I-checks and complement clause and vocabulary probes. Linear models were used for passage-retelling outcomes. Maximum likelihood estimation accounted for missing (Singer & Willett, 2003) or non-binary (Edwards et al., 2012) scores, specifically unscored items in the complement clause tests. There were minimal missing data in the grammar pre-tests (<2%) and non-binary scores on approximately 6% of pre- and post-test scores. Models including random slopes were tested but did not converge and thus only models including random intercepts and fixed slopes are reported in Table 4.

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

In addition to the pre-registered analyses, which focused on group differences, we had questions about the value of the at-home prerecorded books in relation to learning language targets. We noted substantial variability in the degree to which families read these books. Thus, we also conducted post-hoc correlational analyses regarding the contribution of the books to the outcomes. A Pearson correlation was used for the science outcomes because all children were included together, and the data set was large enough to meet assumptions of normality. The grammar and vocabulary results were analysed by arm, and thus a Spearman correlation was used due to the small sample sizes.

Results

Treatment Outcomes

Descriptive statistics appear in Table 3 and are visualised in Figure 2. Estimates of the random and fixed effects for the preregistered model can be found in Table 4 with the science-only condition used as the reference variable. Additional models, with the reference variable shifted to allow for additional pair-wise group comparisons are available in the supplemental materials.

Did the children learn science?

Science pre-test scores did not differ by arm. There was a main effect of time regardless of reference variable, suggesting improvement between pre- and post-testing for all conditions (p-values < .02). Neither arm nor the interaction between arm and time were significant (p > .21), indicating that children learned science content equally well in all three arms, as predicted. Although the gains were modest, they were generally consistent across the three groups, with most children responding to the science instruction (Figure 2). Gains in science outcomes (post-

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

test minus pretest) were not correlated with the number of video log-ins at home for any condition, r(31) = -.20, p =.28)

Did the children learn language targets?

With regard to learning the grammatical targets, there was a main effect of time, regardless of reference variable, such that we observed significant gains from pre- to post-test (p-values < .02) for all groups. Examination of the interactions indicated that gains were largest for the vocabulary arm, which differed from the science-only arm (p = .04) but not the grammar arm (p = .79). The grammar and science-only arms did not differ from each other (p = .16, see Supplemental Materials, Table S1). Grammar gains on the complement clause task did not correlate with the number of video log-ins at home for any condition, Grammar: rho = -.06, p = .86; Vocabulary: rho = .37, p = .32; Science-only: rho = .32, p = .28.

Turning to the changes in taught vocabulary words, only the vocabulary arm showed a significant effect of time (p = .007). The vocabulary arm differed from the other two groups (p-values < .04). The other groups did not show change over time (p > .99), nor did they differ from each other (p > .99). Examination of change in Figure 2 suggests that children in the science-only and grammar arms performed near chance at both pre- and post-test, but the Vocabulary arm performed above chance at post-test. Vocabulary gains did not correlate with the number of video log-ins at home, Grammar: rho = -.45, p = .21; Vocabulary: rho = .23, p =.56; Science-only: rho = .25, p = .40.

Did children generalise science and language gains?

The expression of thematic units and content words in the science retells revealed no effects of time (thematic units p > .12; content words p > .23) or arm (thematic units p > .75; content words p > .21). Time x arm interactions also failed to reach conventional levels of

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

significance (Table 5). Although the children who received vocabulary or grammar intervention did not show overall improvements in science retell, we further examined the data to test the prediction that familiar topics (topics covered in the science lessons) facilitated more complete science retells. That was not the case (Table 2, and Supplemental Materials). In fact, the only hint of improvement was that children in the science + grammar arm used more content words for untaught topics at posttest.

Discussion

We attempted to replicate findings from a prior small-scale randomised controlled trial of in-person language intervention combined with science instruction in a shorter telehealth adaptation that used pre-recorded videos of book reading as an adjuvant to therapy delivered in three arms: science-only control, science + grammar treatment, science + vocabulary treatment. Participants in all arms made gains in science. Only children in the science + vocabulary arm showed gains in vocabulary, despite evidence that they did not fully participate in the book reading adjuvant. Contrary to prediction, all groups learned the grammar targets, with the science + vocabulary arm performing similarly to the science + grammar arm and better than the science-only arm.

Science Learning

The finding that science knowledge improved as much in the science + language arms as in the science-only arm is consistent with other studies that paired science and language instruction in person for typical children (e.g., Hurst et al., 2019; Menninga et al., 2017. It also replicates our findings from in person work with children with DLD (McGregor et al., 2021; Owen Van Horne et al., 2023). Furthermore, the current study demonstrates that inquiry-based

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

science instruction can be delivered via teletherapy, a powerful proof of concept given the ever-increasing use of remote learning (Cary & Ome, 2021).

We hypothesise that three elements were critical to the positive science outcomes. One was having skilled clinicians who were highly familiar with the particular pre-scripted lessons. A second was the students' access to the materials needed to engage in the planned activities. By mailing resource kits to each home, we ensured equitable access for participating families at no expense. Finally, we cannot downplay the caregivers' role in the process, from providing the technology to managing materials to supporting student attention and engagement. Though our approach is promising, it requires careful consideration of instructor preparation, family structure, home life, and access to learning materials and technology to be scalable.

Grammar Learning

The results show that there was growth in complement clause use in all of the treatment arms, which contradicted our hypothesis that the most growth would be seen in the science + grammar arm. The science + vocabulary arm demonstrated more growth than the science-only arm, but NOT the science + grammar arm. That the children improved in all three arms is, perhaps, not so surprising given that science talk offers rich opportunities for complex language use (Curran et al., 2019; Silander et al., 2019). However, in Owen Van Horne et al. (2023), only children in the grammar arm improved on grammar over and above the other two arms that focused on vocabulary and phonological awareness instruction. Thus, it is worth asking what led to gains in this version of the study that were not observed in the summer camp version.

Vocabulary Learning

As predicted, children who received vocabulary instruction improved on the science vocabulary targets; those in the non-vocabulary arms did not. The gains in vocabulary were more

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

robust than those reported for the in-person version of the science + vocabulary intervention (McGregor et al., 2021), but, critically, the previous report reflected a smaller sample, longer sessions, preservice- rather than experienced SLPs, and a focus on word meaning but not word form. In the current study, we maintained the focus on meaning, remaining faithful to the Robust Vocabulary Instruction methods of Beck et al. (2013). However, we added a focus on word form precisely because word form complexity was related to outcomes in McGregor et al. (2021) and because word forms tend to be more challenging than meanings for individuals with DLD in the early stages of learning (Pomper et al., 2022). Thus, we hypothesise that this enhancement to Robust Vocabulary Instruction was critical for these learners.

We note that all three arms were well-matched at pre-test on all collected measures, including nonverbal IQ, receptive and expressive vocabulary, and production probes. That said, they performed better than other groups at both pre- and post- test in expression of science themes during the retell task. Another possibility was that the adjuvant read alouds were incorrectly assigned to participants, but we easily verified that this was not the case. It is also unlikely that the read alouds boosted growth in complement clauses in the vocabulary arm because they were the least compliant with the adjuvant and outcomes were not correlated with use of the read alouds.

Learning Summary

A key difference between Owen Van Horne et al., (2023) and this study was with the way that providers were assigned to intervention. In the in-person version of the study, providers were preservice SLP students supervised by fully credentialed clinicians with at least three years of experience. Students and clinical supervisors were assigned to a single treatment approach to learn and both the clinical supervisors and the live fidelity checkers discouraged use of

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

techniques assigned to other classrooms. In this study the interventionists were fully credentialed SLPs who rotated across arms based on schedule availability and random assignment of groups. Thus, an SLP might move from a science + grammar arm to a science + vocabulary arm in the same day for different groups of children. A post-hoc review of fidelity showed that, in the science + vocabulary and science-only arms, the children heard an average of 13-15 complement clause models per session. Although this is about half the rate of models in the grammar arm, this steady rate of models may have been sufficient to facilitate growth.

We do not view the complement clause models in the non-grammar arm as a failure of experimental control. It is unnatural to teach science without ever using a clausal complement (consider the necessity of sentences like "*we predicted that it would vibrate*"). However, it is similarly odd to teach science without ever using a science-related vocabulary word. The active ingredients 'package' of the grammar arm—elicitations, direct instruction, visual supports, and recasts as well as models—was never used in the non-grammar arms, just as rich instruction plus a focus on word form did not occur in the non-vocabulary arms.

The uniform growth in complement clause use suggests several possibilities that are worthy of further exploration and that may increase the overall practical value of this work. We believe there is a more plausible explanation than maturation, Based on our experience with the targeted nature of grammar intervention in children with DLD, we believe that maturation/typical language development alone would not explain the results of the study. That is, children in the science-only and science + vocabulary arms benefited from an increased rate of exposure (relative to everyday activities) to complement clauses in the absence of explicit instruction or the use of recasting techniques. Our own prior work showed change only in the complement clause groups because the other conditions were actively discouraged from using complement

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

clauses during instruction – thus the additional models provided here is unintentional but differ from prior work and is worthy of further exploration. Modeling, rather than recasting, prompting for production, or explicit didactic teaching techniques, is more feasible to implement at a large-group or classroom level, thus, this is a technique that could be scaled. Asking teachers to use cognitive or communication verbs that naturally obligate complement clauses (e.g., *we thought it would float; they said the wax will melt*) without the need for specialised instructional techniques could thus be effective (Owen Van Horne, et al., 2017). Such an approach is much more feasible than providing ~10 hours of recast therapy focused on a single target.

The doses provided in the science-only and science + vocabulary arms (~13 models per session) were much lower than the dose provided in the science + grammar arm (30 models per session + 20 models per read aloud) yet yielded similar benefits. This suggests that children can learn to produce these types of grammatical forms with a boosted, but not highly- intense, rate of exposure to the target form distributed over time. Dose-finding work, akin to that conducted for word learning (Storkel et al., 2019) is a critical next step for improving the feasibility of this type of implicit therapy in school-based settings.

Furthermore, the finding that children improved on both grammar and vocabulary in the science + vocabulary arm is encouraging. Prior work on combining targets is limited. Our work contradicts Tyler et al. (2002) who concluded that intervention should only focus on a single target at a time. In that case, the focus was on the co-presentation of phonological and morphological targets (e.g., final consonant deletion or cluster reduction and use of past tense – ed). We are unaware of studies that focus on both vocabulary and syntax. It is worth noting that some Narrative Language Intervention programmes, like Story Champs, attempt to target both vocabulary and syntax, alongside story grammar. However, the ability of these programmes to

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

teach syntax to children with DLD is unclear. For instance, in a review of their own work, Petersen & Spencer (2016) report successfully improving syntax in typical children but not children with DLD in the absence of direct instruction focused on the grammatical forms. In our study here, we show that vocabulary words were learnable coincident with complement clauses and science content. This suggests that when instruction techniques are systematic and high quality and provided in a repeated fashion over time, children with DLD can benefit from exposure to multiple types of targets simultaneously. Further exploration about the dose and degree of direct and indirect instruction in each domain is required.

Role of at home reading as an adjuvant

Although we carefully crafted and recorded books to augment both science and language content, we do not have demonstrated evidence that these pre-recorded books had benefit in any domain. There was no correlation between rate of book reading and degree of benefit. It may be that the books were not sufficiently interesting for the children to engage with and attend to, since young children benefit more from live presentations (Kim, 2016). It may also be that children learned as much as they were likely to learn in that time frame from the interactive elements, and while an effect of book reading *in isolation* might have been seen, book reading *as an adjuvant* did not provide added benefit. Further investigation is worthwhile as book reading to support language outcomes is well-established (e.g., Goldstein et al., 2017).

Limitations and Future Directions

The external validity of this study is limited by the small sample size, the focus on preschoolers and the science curriculum, and the need to include only families who already had access to technology and who had the time to supervise and support their child's participation at home. Additional work to consider the role of the family in the success of this type of teletherapy

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

approach and to extend our findings to a larger sample, other academic domains, and older grade levels is warranted, as well as rigorous testing in a trial with some children receiving no intervention. For instance, one might predict that comparative terms or conditionals are better taught in a mathematical context. The lack of a true control group also limits the conclusions that we can draw since all groups made similar progress.

Another concern is the minimal generalisation effects as measured by distal tools. Story length as measured by number of content words and number of thematic units did not improve with intervention and was not better when those stories cover the topics we taught than when they covered new topics. It may be that short-term programs like this do not lead to functional change, or it may be that the distal measure is not sensitive to small amounts of change. We suspect that we have not yet identified sufficiently sensitive measures. In our prior work, neither retells of Sid the Science Kid engineering-focused episodes nor a computer-adaptive science assessment were effective distal measures. Each had their own challenges that affected sensitivity. For instance, the computer-adaptive science assessment required children to demonstrate proficiency in following directions prior to the science assessment and this led to a large amount of missing data. In this case, the retells, though we had several, were short in nature. Spontaneous retells of science-related topics are often too short to show reliable effects of change in language skills at this age. In fact, most national English language arts and science assessments begin at 3rd grade. Thus, a critical step required to advance this line of research is the development of measures sensitive to language and science-related curricular changes at these younger ages.

Conclusion

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

This study yields three primary conclusions. Children with DLD can learn language targets in the context of curricular instruction; thus, SLPs should ensure that treatment is functional and tied to academic outcomes to provide maximum benefit. Children with DLD can benefit from telehealth methods of instruction for curricular content and language targets; thus, SLPs can offer self-service delivery solutions for rural or other hard-to-reach populations or families who find it more convenient and can support engagement. Finally, Rich Vocabulary Instruction enhanced with input that emphasises word forms holds promise as an approach that can be embedded in the curriculum and produces broad gains; thus, SLPs might consider partnering with teachers to implement this approach in the classroom setting.

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

References

Archibald, L. M. (2017). SLP-educator classroom collaboration: A review to inform reason-based practice. *Autism & Developmental Language Impairments*, *2*, 2396941516680369. https://doi.org/10.1177/2396941516680369

Arnold, H. L., Plante, E., & Vance, R. (2022). Translating enhanced conversational recast to a telepractice setting. *Language, Speech, and Hearing Services in Schools*, *53*(2), 275-289. doi: 10.1044/2021_LSHSS-21-00051.

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press. https://www.guilford.com/books/Bringing-Words-to-Life/Beck-McKeown-Kucan/9781462508167

Byers-Heinlein, K., Schott, E., Gonzalez-Barrero, A. M., Brouillard, M., Dubé, D., Jardak, A., ... & Tamayo, M. P. (2020). MAPLE: A multilingual approach to parent language estimates. *Bilingualism: Language and Cognition*, *23*(5), 951-957. DOI 10.17605/OSF.IO/BYXFZ

Campbell, D. R., & Goldstein, H. (2022). Evolution of telehealth technology, evaluations, and therapy: Effects of the COVID-19 pandemic on pediatric speech-language pathology services. *American Journal of Speech-Language Pathology*, *31*(1), 271-286. doi: 10.1044/2021_AJSLP-21-00069.

Castilla-Earls, A., & Owen Van Horne, A. (2023). Recast therapy for treating syntax in bilingual children with developmental language disorder: A feasibility and early efficacy study examining the role of language of intervention on outcomes. *Journal of Speech, Language, and Hearing Research*, *66*(8), 2783-2801. doi: 10.1044/2023_JSLHR-22-00452.

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Curran, M., & Owen Van Horne, A. (2019). Use of recast intervention to teach causal adverbials to young children with developmental language disorder within a science curriculum: A single case design study. *American Journal of Speech-Language Pathology*, *28*(2), 430-447. doi: 10.1044/2018_AJSLP-17-0164.

Dillon, H., Mee, C., Moreno, J. C., & Seymour, J. (2018). Hearing tests are just child's play: the sound scouts game for children entering school. *International Journal of Audiology*, *57*(7), 529-537. doi: 10.1080/14992027.2018.1463464

Duff, D. M., Hendricks, A. E., Fitton, L., & Adlof, S. M. (2022). Reading and math achievement in children with dyslexia, developmental language disorder, or typical development: Achievement gaps persist from second through fourth grades. *Journal of Learning Disabilities*, 00222194221105515. doi: 10.1177/00222194221105515.

Dunn, D. M. (2019). Peabody Picture Vocabulary Test (5th ed.) [Measurement instrument]. Bloomington, MN: NCS Pearson. https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Academic-Learning/Peabody-Picture-Vocabulary-Test-%7C-Fifth-Edition/p/100001984.html

Edwards, J. K., Cole, S. R., Troester, M. A., & Richardson, D. B. (2013). Accounting for misclassified outcomes in binary regression models using multiple imputation with internal validation data. *American Journal of Epidemiology, 177*(9), 904-912. doi: 10.1093/aje/kws340.

Elliott, C. D., Salerno, J. D., Dumont, R., & Willis, J. O. (2007). *Differential ability scales, Second edition.* San Antonio, TX:Harcourt Assessment https://doi.org/10.1177/0734282910368783

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Finestack, L. H., & Satterlund, K. E. (2018). Current practice of child grammar intervention: A survey of speech-language pathologists. *American Journal of Speech-Language Pathology*, *27*(4), 1329-1351. doi: 10.1044/2018_AJSLP-17-0168

Goldstein, H., Ziolkowski, R. A., Bojczyk, K. E., Marty, A., Schneider, N., Harpring, J., & Haring, C. D. (2017). Academic vocabulary learning in first through third grade in low-income schools: Effects of automated supplemental instruction. *Journal of Speech, Language, and Hearing Research*, *60*, 3237-3258. doi: https://doi.org/10.1044/2017_JSLHR-L-17-0100

Haden, C. A., Jant, E. A., Hoffman, P. C., Marcus, M., Geddes, J. R., & Gaskins, S. (2014). Supporting family conversation and children's STEM learning in a children's museum. *Early Childhood Research Quarterly, 29*, 333-344. https://doi.org/10.1016/j.ecresq.2014.04.004

Hurst, M. A., Polinsky, N., Haden, C. A., Levine, S. C., & Uttal, D. H. (2019). Leveraging research on informal learning to inform policy on promoting early STEM. Society for Research in Child Development. *Social Policy Report, 32*(3), 1-33. https://doi.org/10.1002/sop2.5

Kim, Y.-S., G. (2015). Do live versus audio-recorded narrative stimuli influence young children's narrative comprehension and retell quality? *Language, Speech, and Hearing Services in Schools*, *47*, 77-86. doi: 10.1044/2015_LSHSS-15-0027

The Lawrence Hall of Science. (2015). Foss Full Option Science System. Retrieved August 13, 2023, from https://foss.lawrencehallofscience.org.

Mantzicopoulos, P., & Patrick, H. (2010). "The seesaw is a machine that goes up and down": Young children's narrative responses to science-related informational text. *Early Education and Development*, *21*(3), 412-444.

McGregor, K. K., Van Horne, A. O., Curran, M., Cook, S. W., & Cole, R. (2021). The challenge of rich vocabulary instruction for children with developmental language disorder. *Language, Speech, and Hearing Services in Schools*, *52*(2), 467-484. doi: 10.1044/2020_LSHSS-20-00110.

Mealings, K., Harkus, S., Flesher, B., Meyer, A., Chung, K., & Dillon, H. (2020). Detection of hearing problems in Aboriginal and Torres strait islander children: a comparison between clinician-administered and self-administrated hearing tests. *International Journal of Audiology*, *59*(6), 455-463. doi: 10.1080/14992027.2020.1718781.

Meier, L. (2022). ANOVA and Mixed Models: A Short Introduction Using R (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003146216.

Menninga, A., van Dijk, M., Steenbeek, H., & van Geert, P. (2017a). Language use in real-time interactions during early elementary science lessons: The bidirectional dynamics of the language complexity of teachers and students. *Language Learning, 67*(2), 284-320. https://doi.org/10.1111/lang.12219

Menninga, A., van Dijk, M., Wetzels, A., Steenbeek, H., & van Geert, P. (2017b). Language use in kindergarten science lessons: Language production and academic language during a video feedback coaching intervention in kindergarten science lessons. *Educational Research and Evaluation, 23*(1), 26-51. doi: 10.1080/13803611.2017.1292920

Owen Van Horne, A. J., Curran, M., Cook, S. W., Cole, R., & McGregor, K. K. (2023). Teaching little kids big sentences: A randomized controlled trial showing that children with DLD

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

respond to complex syntax intervention embedded within the context of preschool/kindergarten science instruction. *International Journal of Language & Communication Disorders*. doi: 10.1111/1460-6984.12882.

Peña, E. D., & Sutherland, R. (2022). Can you see my screen? Virtual assessment in speech and language. *Language, Speech, and Hearing Services in Schools*, *53*(2), 329-334. https://doi.org/10.1044/2022_LSHSS-22-00007

Petersen, D. B., & Spencer, T. D. (2016). Using narrative intervention to accelerate canonical story grammar and complex language growth in culturally diverse preschoolers. *Topics in Language Disorders, 36*(1), 6-19. doi:10.1097/TLD.0000000000000078

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, *25,* 15-24. https://doi.org/10.1044/0161-1461.2501.15

Pomper, R., McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Ohlmann, N. (2022). Direct instruction improves word learning for children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *65*(11), 4228-4249. doi: 10.1044/2022_JSLHR-22-00300.

Pratt, A. S., Anaya, J. B., Ramos, M. N., Pham, G., Muñoz, M., Bedore, L. M., & Peña, E. D. (2022). From a distance: Comparison of in-person and virtual assessments with adult–child dyads from linguistically diverse backgrounds. *Language, Speech, and Hearing Services in Schools*, *53*(2), 360-375. doi: 10.1044/2021_LSHSS-21-00070.

Restrepo, M. A., Morgan, G. P., & Thompson, M. S. (2013). The efficacy of vocabulary intervention for dual-language learners with language impairment. *Journal of Speech,*

Language, and Hearing Research, 56, 748-765. doi: 10.1044/1092-4388(2012/11-0173)x.

Schmitt, M. B., Tambyraja, S., Thibodeaux, M., & Filipkowski, J. (2022). Feasibility of assessing expressive and receptive vocabulary via telepractice for early elementary-age children with language impairment. *Language, Speech, and Hearing Services in Schools*, *53*(2), 445-453. doi: 10.1044/2021_LSHSS-21-00057.

Sealed Envelope Ltd. 2020. Create a blocked randomisation list. Available from: https://www.sealedenvelope.com/simple-randomiser/v1/lists [Accessed 5 Feb 2021].

Silander, M., Cerrone, M., Cuellar, L., Hiebert, L., & Stiles, J. (2019). Bringing science home with PEEP: Using digital media to foster family engagement in science at home. Executive Summary and Final Report [Grant No. 1612643]. WGBH Educational Foundation. https://cct.edc.org/sites/cct.edc.org/files/publications/PEEP7FinalReport.pdf

Singer, J. D., & Willet, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford University Press, Inc. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

Steel, G., Rose, M., & Eadie, P. (2016). The production of complement clauses in children with language impairment. *Journal of Speech, Language, and Hearing Research*, *59*(2), 330-341. doi: 10.1044/2015_JSLHR-L-15-0001.

Storkel, K. L., Komesidou, R., Pezold, M. J., Pitt, A. R., Fleming, K. K., & Romine, R. S. (2019). The impact of dose and dose frequency on word learning by kindergarten children with developmental language disorder during interactive book reading. *Language, Speech, and Hearing Services in Schools*, *50,* 518-539. doi: 10.1044/2019_LSHSS-VOIA-18-0131.

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Strouse, G. A., O'Doherty, K., & Troseth, G. L. (2013). Effective coviewing: Preschoolers' learning from video after a dialogic questioning intervention. Developmental Psychology, 49(12), 2368–2382. doi: 10.1037/a0032463.

Sutherland, R., Hodge, A., Chan, E., & Silove, N. (2021). Barriers and facilitators: Clinicians' opinions and experiences of telehealth before and after their use of a telehealth platform for child language assessment. *International Journal of Language & Communication Disorders*, *56*(6), 1263-1277. doi: 10.1111/1460-6984.12666.

Sutherland, R., Hodge, A., Trembath, D., Drevensek, S., & Roberts, J. (2016). Overcoming barriers to using telehealth for standardized language assessments. *Perspectives of the ASHA Special Interest Groups*, *1*(18), 41-50. https://doi.org/10.1044/persp1.SIG18.41

Sutherland, R., Trembath, D., Hodge, A., Drevensek, S., Lee, S., Silove, N., & Roberts, J. (2017). Telehealth language assessments using consumer grade equipment in rural and urban settings: Feasible, reliable and well tolerated. *Journal of Telemedicine and Telecare*, *23*(1), 106-115. doi: 10.1177/1357633X15623921.

Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P., & Landis, J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics, 54*(1), 367-383.

Tyler, A. A., Lewis, K. E., Haskill, A., & Tolbert, L. C. (2002). Efficacy and Cross-Domain Effects of a Morphosyntax and a Phonology Intervention. *Language, Speech, and Hearing Services in Schools*, *33*(1), 52-66. doi: 10.1044/0161-1461(2002/005).

Williams, K. T. (2019). *Expressive Vocabulary Test* (3rd ed.) [Measurement instrument]. Bloomington, MN: NCS Pearson.

Zhang, Z. & Wang, L. (2017-2022). *Advanced statistics using R*. Granger, IN: ISDSA Press. https://doi.org/10.35566/advstats. ISBN: 978-1-946728-01-2.

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Table 1. Demographic information and descriptive statistics for participants and caregivers in each intervention group.

	Science Group	Grammar Group	Vocabulary Group
Child Age	4.98 (1.07)	4.89 (1.24)	4.89 (1.10)
Parent Education	16.07 (2.75)	16.35 (2.55)	16.04 (2.62)
DELV Standard Score	78.87 (5.12)	76.26 (5.49)	76.54 (5.17)
DAS T-Score	49.52 (9.67)	49.89 (11.26)	49.82 (10.02)
PPVT Standard Score	90.42 (8.75)	89.11 (9.31)	90.14 (8.93)
EVT Standard Score	91.16 (7.69)	90.95 (7.46)	91.21 (8.08)
	Sex		
Female	3	5	2
Male	10	5	7
	Race		
White/Caucasian	9	5	7
Black/African American	2	3	1
Asian/Pacific Islander	2	1	1
American Indian /Alaska Native	0	1	0
	Ethnicity		
Hispanic	1	1	1
Non-Hispanic	12	9	8

Note. DELV = Diagnostic Evaluation of Language Variation-Norm Referenced; DAS = Differential Abilities Scale-II; PPVT = Peabody Picture Vocabulary 4th Edition Test; EVT = Expressive Vocabulary Test 3rd Edition.

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Table 2. Average fidelity and cross contamination of concepts taught across science, grammar, and vocabulary arms for sessions reviewed.

	Science-Only		Science+Grammar		Science+Vocabulary	
	Planned	Taught	Planned	Taught	Planned	Taught
Science Concepts	16-34	21.4 (6.7)	16-34	21.8 (6.2)	16-34	22.2 (6.6)
Grammar Models	0	**12.9 (6.3)**	30-45	78.5 (19.6)	0	**14.5 (5.4)**
Grammar Visual Cue	0	0	3	8	0	0
Grammar Definition	0	0	1	1.8	0	0
Grammar Elicitation	0	0	3	3.2	0	0
Grammar Recast	0	0	10-15	13.1 (5.6)	0	0
Vocabulary Models	0	**4.8 (7.8)**	0	**2.4 (2.7)**	12-18	19.2 (7.4)
Vocabulary Definition	0	0	0	0	4	2.6 (1.9)
Vocabulary Sound Cue	0	0	0	0	4	3.0 (2.3)
Vocabulary Elicitation	0	0	0	0	8-12	3.3 (1.6)

Note. Bolded values differ from expected in ways that suggest elements of another arm were present at high rates. Number ranges reflect different numbers of participants per group (either two or three).

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Table 3. Means and standard deviations for assessment scores, and pre- and post-test percentage scores by intervention group.

Outcomes	Test	Time	Science Group	Grammar Group	Vocabulary Group
Proximal[a]	Science Unit I-Check	Pre	43.01 (9.60)	41.40 (9.78)	42.62 (9.74)
		Post	53.23 (12.34)	50.88 (12.36)	52.02 (12.61)
	Grammar Probes	Pre	8.87 (11.35)	8.16 (9.79)	6.88 (1.00)
		Post	20.60 (20.06)	19.74 (17.89)	18.53 (19.89)
	Vocabulary Probes	Pre	50.00 (18.32)	45.83 (13.75)	49.07 (19.74)
		Post	51.92 (21.01)	45.00 (17.21)	63.89 (21.25)
Distal	Target Science Topics				
	Content Words[b]	Pre	2.27 (1.77)	2.97 (1.64)	2.38 (1.75)
		Post	3.00 (1.76)	2.73 (1.23)	2.85 (2.41)
	Thematic units[c]	Pre	1.33 (.96)	1.43 (.86)	1.22 (.93)
		Post	1.69 (1.08)	1.90 (1.06)	1.81 (1.30)
	Non-target Science Topics				
	Content Words	Pre	2.64 (1.66)	2.83 (1.63)	3.04 (1.99)
		Post	2.77 (1.63)	3.17 (1.80)	3.26 (2.10)
	Thematic units	Pre	1.59 (1.07)	1.77 (.97)	2.04 (0.98)
		Post	1.97 (1.20)	1.90 (.96)	2.52 (1.01

Note. [a]Scores = percent correct; [b]Scores = number of content words in the stories that were included in the retellings (maximum possible ranged from 14 to 17 across stories); [c]Scores = Thematic units were scored using a prepared rubric and had a potential range of 1 to 4.

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Table 4. Logistic mixed effects model for change in participants' pre- and post-test scores for science, grammar, and vocabulary assessments.

Science Assessment Scores				
Random effects	**s^2**	***SD***		
Participant (Intercept)	0.14	0.38		
Item (Intercept)	1.91	1.38		
Fixed effects	**b**	***SE***	***z***	***p***
(Intercept)				
Ref = Science-Only+Pretest	-0.41	0.30	-1.36	0.18
Posttest	0.40	0.17	2.37	0.02*
Grammar	-0.18	0.24	-0.73	0.46
Vocabulary	0.08	0.25	0.30	0.76
Posttest:Grammar	0.21	0.26	0.83	0.41
Posttest:Vocabulary	0.33	0.26	1.25	0.21
Grammar (Complement Clause) Assessment Scores				
Random effects	**s^2**	***SD***		
Participant (Intercept)	2.48	1.57		
Item (Intercept)	0.03	0.16		
Fixed effects	**b**	***SE***	***z***	***p***
(Intercept)				
Ref = Science-Only+Pretest	-3.37	.57	-5.88	<0.001***
Posttest	.77	.32	2.43	.02*
Grammar	-.68	.87	-.78	.44
Vocabulary	.12	.84	.15	.88
Posttest:Grammar	.81	.57	1.41	.16
Posttest:Vocabulary	.96	.48	2.01	.04*
Vocabulary (Taught Only) Assessment Scores				
Random effects	**s^2**	***SD***		
Participant (Intercept)	0.35	0.59		
Item (Intercept)	0.45	0.67		
Fixed effects	**b**	***SE***	***z***	***p***
(Intercept)				
Ref = Science-Only+Pretest	-0.04	0.35	-0.14	0.88
Posttest	0.00002	.25	0.00	.99
Grammar	-0.28	.37	-.76	0.45
Vocabulary	-0.15	.38	-.40	0.69
Posttest:Grammar	0.00008	.38	0.00	0.99
Posttest:Vocabulary	.84	.40	.21	0.04*

Note. s^2 = Variance; *SD* = Standard Deviation; b = Estimate; *SE* = Standard Error; *z* = z value; *p* = significance. *p < .05. ***p < .001.

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Table 5. Linear mixed effects models for change in distal retells from pre- to post-test.

Rubric for Science Themes					
Random effects	**s^2**	***SD***			
Participant (Intercept)	.42	.65			
Subtest (Intercept)	.02	.13			
Residual	.67	.82			
Fixed effects	**b**	***SE***	***df***	***t-value***	***p***
(Intercept)					
Ref = Science-Only+Pretest	1.33	.23	31.15	5.67	<.0001**
Posttest	.36	.21	12.82	1.68	.12
Grammar	.10	.34	42.06	.29	.77
Vocabulary	-.11	.35	42.06	-.31	.75
Posttest:Grammar	.11	.28	153	.38	.70
Posttest:Vocabulary	.23	.29	153	.80	.42
Science-related Content Words					
Random effects	**s^2**	***SD***			
Participant (Intercept)	1.09	1.04			
Item (Intercept)	.30	.55			
Residual	1.96	1.39			
Fixed effects	**b**	***SE***	***df***	***t-value***	***p***
(Intercept)					
Ref = Science-Only+Pretest	2.23	.48	13.81	4.59	<.0001**
Posttest	.73	.55	6.31	1.33	.23
Grammar	.70	.56	44.19	1.25	.21
Vocabulary	.15	.58	44.25	.25	.80
Posttest:Grammar	-.93	.48	148.31	-1.93	.06
Posttest:Vocabulary	-.26	.49	148.32	-.52	.60

Note. s^2 = Variance; *SD* = Standard Deviation; b = Estimate; *SE* = Standard Error; *df* = degrees of freedom; *p* = significance. ***p < .001

Accepted Manuscript
Version of Record at: https://doi.org/10.1080/17549507.2024.2445155

Figure Captions

Figure 1. Consort flow diagram of families recruited and participants who were withdrawn, consented, and participated.

Figure 2. Gains on Science, Grammar, and Vocabulary measures for each Treatment Arm.

Supplementary Materials