An individualized multimodal narrative intervention for young children with neurodevelopmental disorders

Júlia Florit-Pons, Pilar Prieto, Alfonso Igualada

Abstract

This study examines whether an individualized narrative intervention (*MultiModal Narrative*) can boost the narrative (macrostructure and perspective-taking) abilities of children with neurodevelopmental disorders (NDD). An experimental group of 16 children with NDD from Catalonia, Spain, received 9 MultiModal Narrative intervention sessions. Children's narrative skills were measured pre- and post-intervention with a retelling task. Their performance was compared to two control groups not receiving the intervention (17 children with NDD and 17 with typical development). After each session, children in the experimental group underwent two Dynamic Assessment measures (retelling task and graduated prompting comprehension questions about the story), tracking their learning. Results showed an improvement in narrative macrostructure but not perspective-taking. The Dynamic Assessment revealed faster learning of macrostructure, individual differences in the prompts needed for successful responses, and that fewer prompts predicted better macrostructure outcomes post-intervention. These findings underscore the relevance of diverse measures in evaluating continuous learning in interventions.

Keywords: narrative intervention, multimodality, neurodevelopmental disorders, narrative macrostructure skills, narrative perspective-taking skills, dynamic assessment

Introduction

The ability to construct an oral narrative discourse by retelling or generating a story is one of the key milestones in children's language development, as it constitutes a complex linguistic ability requiring not only the use of the appropriate vocabulary and morphosyntactic structures but also organizing these structures in the correct order while framing them in a relevant socio-communicative context to express what happened in the story (i.e., narrative macrostructure) and what were the perspectives and emotions of the characters (i.e., narrative perspective-taking). These narrative skills have been shown to be both predictors and precursors of later language skills, academic success, and socialization (e.g., Babayiğit et al., 2021; Dickinson & McCabe, 2001). It is during the ages of 4 to 6 that children start producing narrative discourses that are more complex and include the core elements of a story, structured coherently and cohesively. Nevertheless, children with neurodevelopmental disorders (NDD), such as autism or Developmental Language Disorder, overlap in their difficulties with narrative and socio-communicative skills (see Astle et al., 2022, proposing a transdiagnostic approach). Overall, children with NDD (both autism and Developmental Language Disorder) have been shown to produce narrative discourses that are significantly below those of age-matched typically developing peers (e.g., Acosta-Rodríguez et al., 2022; Baixauli et al., 2016; García-Pérez et al., 2008; Norbury & Bishop, 2003; Winters et al., 2022). It has been shown by recent review studies (e.g., Acosta-Rodríguez et al., 2022; Baixauli et al., 2016; García-Pérez et al., 2008; Norbury & Bishop, 2003; Winters et al., 2022) that young children with NDD (both children with autism and Developmental Language Disorder) struggle with narrative macrostructure skills, such that they have difficulties in identifying the different core elements of a story, which diminishes their capacity to build a complex and coherent narrative discourse. As for narrative perspective-taking, there is evidence indicating that the narrative discourses of both children with autism and Developmental Language Disorder tend to include minimal or no reference to

characters' emotions and perspectives (see e.g., Baixauli et al., 2016; Norbury et al., 2014; Winters et al., 2022).

To foster oral narrative skills in children with NDD, many researchers and practitioners have been devoted to designing and implementing narrative-based interventions for this population. As reported by a recent systematic literature review examining 24 narrative intervention studies for children with language disorder aged between 3 and 18 (Favot et al., 2021), most interventions have been developed to target preschool and school-aged children's narrative macrostructure skills (i.e., the organization of the different narrative events to express what happened in the story) and/or linguistic productivity or complexity within the discourse (also known as narrative microstructure). Although studies have reported the beneficial effects of narrative instruction on both macrostructure and linguistic complexity and productivity (with moderate to large effect sizes), the effect sizes seem to be larger for macrostructure, suggesting that it is easier for children to improve in terms of identifying and reproducing the story events. An important factor in intervention research is the delivery of the interventions. Particularly for children with NDD, following the multi-tiered system of support guidelines (e.g., Clark & Dockweiler, 2019; Jimerson et al., 2016), interventions are usually implemented at an individual (or small group) level with more intensive instruction (see also Ireland & Hall-Mills, 2024). In this regard, most narrative-based intervention studies with NDD children have been implemented individually with a set of intensive instruction measures and strategies, such as repetitive activities and adaptable intervention pace, the use of visuals like story icons or pictures or modeling by the interventionist, and direct instruction of the story grammar (see Favot et al., 2021 and Pico et al., 2021 for reviews). All these strategies implemented by speech-language therapists delivering the interventions have been found to serve as effective tools for fostering children's oral narrative skills (see Spencer & Petersen, 2020, for a detailed summary of the instructional principles of narrative intervention). Interestingly, recent studies

have shown that oral narrative instruction also transfers to literacy skills, such as listening, reading, or writing in both children with NDD (see e.g., Whalon et al., 2019, assessing 5- to 7-year-old autistic children) and those with typical development (TD) (Spencer & Petersen, 2018, assessing 6- to 7-year-olds).

To implement effective interventions that adapt to children's needs, speech-language professionals typically emphasize their need to incorporate tools to track children's learning ability and potential for learning throughout the intervention (see blinded, under review, for an assessment of professionals' needs in the context of narrative interventions in Catalonia). One of these tools is Dynamic Assessment, which is an interactive evaluation method that can help practitioners understand an individual's learning needs and identify and predict learning potential (Bamford et al., 2022). In fact, Dynamic Assessment has been found to be effective in predicting children's language difficulties and diagnosing language disorders (e.g., Petersen et al., 2017; for reviews, see Bamford et al., 2022; Orellana et al., 2019). Using Dynamic Assessment measures during the language intervention process can have diverse purposes, such as monitoring changes throughout the intervention, planning or adjusting the intervention delivery, and predicting changes after the intervention (Camilleri & Botting, 2013; Hasson et al., 2012; Olswang & Bain, 1996; for a review, see Bamford et al., 2022). For instance, Hasson and colleagues (2012) showed that the number of prompt cues during Dynamic Assessment (in the context of learning syntactic structures) significantly correlated with long-term standardized core language scores after the intervention. Similarly, Camilleri and Botting (2013) assessed 3- and 4-year-old children who had been referred to speech-language therapy for language difficulties and reported that their need for supporting learning was positively correlated with vocabulary scores at post-test. More particularly, Dynamic Assessment has been implemented within the context of narrative abilities using different procedures, such as test-train-retest or prompts together with testing (e.g., Fiestas & Peña, 2018; Peña et al., 2006;

2014). For instance, Peña et al. (2006), implemented two individualized narrative intervention sessions incorporating Dynamic Assessment in children aged 6 to 7 with language impairment and typical development. They showed that the intervention had small to moderate effect sizes on children's learning after the sessions. In line with this, Fiestas and Peña (2018) reported that typically-developing 6- to 9-year-old children's narrative macrostructural abilities improved after a brief 2-session learning experience in both English and Spanish (with a moderate effect size). Nevertheless, to our knowledge, no previous narrative-based intervention studies have examined the effect of Dynamic Assessment prompting on predicting the change in children with NDD after intervention.

Despite the increasing evidence showing that narrative-based interventions are effective in boosting oral language skills (see Favot et al., 2021 for a review; see also Pico et al., 2021), most narrative interventions have focused on narrative macrostructure and have generally avoided training and assessing narrative perspective-taking skills, which are relevant for understanding how the characters feel and think throughout the story (see Dodd et al., 2011; Hessling & Schuele, 2020 for exceptions, assessing 8- to 12-year-old children with autism and Developmental Language Disorder, respectively). Additionally, although there is a great amount of evidence highlighting the beneficial role of using multimodal strategies (such as gestures or story enactment) to improve language skills (e.g., Bernstein et al., 2024; Dargue et al., 2021; Frey & Lüke, 2023; Nicolopoulou et al., 2015; for reviews, see Rohlfing, 2019; Vilà-Giménez & Prieto, 2021, with moderate to large effect sizes), the vast majority of narrative interventions for children with NDD do not integrate them systematically. Also, there is a lack of narrative-based interventions that systematically incorporate Dynamic Assessment measures, despite their reported effectiveness in monitoring children's learning ability and in predicting changes in learning (e.g., Bamford et al., 2022). Finally, most of them have been designed for English-speaking or English-Spanish bilingual children and have not considered

other languages (with some exceptions, such as Delgado-Cruz et al., 2022 and Delgado-Cruz et al., 2024 for 5-year-old Spanish-speaking children with Developmental Language Disorder), nor have they investigated other contexts. Assessing the effectiveness of oral narrative interventions in other contexts might be of interest, as each context has its own differential characteristics. For instance, in the context of Catalonia, where the current study was conducted, preschool and early school education is mostly focused on early literacy skills, rather than oral discourse, meaning that professionals lack validated materials in the Catalan language that address their needs if they want to intervene in oral narrative abilities. Therefore, assessing other educational and clinical contexts has implications, not only to have research evidence on the generalizability of narrative interventions, but also practical implications for professionals, who will be provided with validated materials in their language that are adapted to their context.

To address all these issues, we developed a multi-tiered multimodal narrative-based intervention, which was co-created with 93 professionals from Catalan clinical and educational contexts (see section *The MultiModal Narrative Intervention* for more details; see also Florit-Pons et al., 2024, in press). The intervention trains both narrative macrostructure and perspective-taking through a set of educational strategies, including interactive and multimodal enactment strategies using Catalan, the vehicular language in schools in Catalonia. The intervention design was couched on two main theoretical frameworks, namely (a) the social interactionist theory, indicating that children learn from the interaction with others in a social context (e.g., Kuhl, 2014; Lytle & Kuhl, 2017); and (b) the multimodal theories of enriched learning (Mathias & von Kriegstein, 2023), suggesting that multimodality can boost language learning. This study aims to assess whether an intensive implementation of this multimodal-based narrative intervention can boost both narrative macrostructure and perspective-taking skills of children with NDD, who have difficulties with oral narrative skills, following a

transdiagnostic approach (Astle et al., 2022; Donolato et al., 2023). To do so, two sets of complementary measures were used, namely static measures comparing pre- and post-intervention oral narrative outcomes and Dynamic Assessment session-by-session measures to assess continuous learning and potential for learning. Crucially, we tested whether individual performance on Dynamic Assessment narrative measures can predict macrostructure gains after the intervention. First, we hypothesize that children's oral narrative skills might improve significantly after receiving an intervention that systematically trains narrative macrostructure and narrative perspective-taking. Second, we expect to observe children's continuous narrative learning process through the Dynamic Assessment measures, which will serve to predict their narrative skills post-intervention.

Methods

To address our aim, we followed a non-randomized control trial design that included three groups (an experimental group with NDD children, a control group with NDD children, and a control group with TD children). Having these three groups allowed for comparing the effectiveness of the intervention at two different levels. First, the comparison with the control NDD group allowed us to differentiate the intervention effects from the general developmental progress and control for task repetition effects. Second, the control TD group served as a baseline to compare NDD children's narrative abilities to those of children with typical development, and thus this helped to assess whether after the intervention, NDD children's performance is similar to that of children with TD.

Participants

Fifty children (N = 50) from an initially recruited sample of 66 participated in the study (see Appendix A for the CONSORT 2010 flow diagram depicting the process from enrollment until analysis). Sample size was calculated *a priori* using G*Power Version 3.1.9.6 (Faul et al., 2009) with inputs of .05 for alpha, 3 groups, and 2 for the number of measurements (pre- and post-intervention). The necessary sample size to have an acceptable level of power (1-β error probability = 0.8) and to detect small effect sizes of .25 was 42 participants.

Participants were TD and NDD children aged 3 to 7 years from Catalonia, Spain (see Table 1 for participants' characteristics and Appendix B for the distribution of participants within the NDD groups). All the participants were Spanish-Catalan bilinguals. All children were enrolled in public schools and were therefore taught in a general education setting. TD participants were recruited from a public school in a city in the metropolitan area of Barcelona. As NDD were receiving additional intervention services apart from the general education service at school, they were recruited through contact with speech-language therapists working at public and private speech-language therapy services, or with language-specialized teachers working at public schools. NDD children were required to meet the following inclusion criteria to be eligible for participation in the study: a) have a previous professional diagnosis or reported risk of autism or Developmental Language Disorder (following a transdiagnostic approach, Astle et al., 2022); b) have a vocabulary size of at least 50 functional words; c) be able to use 2-word combinations systematically; and d) receive weekly intervention sessions with a certified specialist.

In this study, a transdiagnostic approach was followed (Astle et al., 2022) because of the following reasons. First, some children did not yet have a differential diagnosis, because they were still young and also because there was a lot of variability in how diagnostic labels were assigned. This is in line with the general reports that young children are often undiagnosed (McGregor, 2020). To solve this issue, some participants were considered to be at risk of being

diagnosed with autism or Developmental Language Disorder. Second, despite not having an official diagnosis, at the time of testing, all children were undergoing assessments and were in the process of receiving a diagnosis. The fact that children with autism and Developmental Language Disorder have been shown to have overlapping difficulties with oral narrative and pragmatic abilities (e.g., Norbury et al., 2014; Norbury & Bishop, 2003) and that there is recent evidence suggesting that oral language interventions have similar aims and effects in these groups of children (see Donolato et al., 2023) also contributed to following a transdiagnostic approach. Finally, we follow CATALISE's (Bishop et al., 2016, 2017) approach to prioritize more the needs children have (i.e., functional diagnosis) rather than a diagnostic label. For this, we ensured that all participating children had a professional report describing discourse-level linguistic difficulties and that their usual speech-language therapists confirmed that a narrative intervention was appropriate for them, regardless of children not being previously exposed to any narrative intervention either at the school level or at the individual speech-therapy level.

Before starting the intervention, participants' core language abilities (either CELF Preschool-2 or CELF-5 core language score, depending on the child's age; Wiig et al., 2009, 2013), pragmatic abilities (PleaseApp; Andrés-Roqueta et al., 2024), non-verbal IQ (K-BIT Matrices; Kaufman & Kaufman, 1990) were evaluated to assess group comparisons using pairwise t-tests (see Table 2 for all the statistical results, with Cohen's *d* as a measure of effect size). Interestingly, no significant differences in core language, receptive pragmatics and non-verbal IQ were found between the participants with autism or Developmental Language Disorder (CELF: $p = .070$; PleaseApp: $p = .129$; K-BIT: $p = .987$).

Ethical approval was obtained from the Institutional Committee for Ethical Review of Projects at the host university of the first and second authors (ref.: 228) and by the Regional Ministry of Education (Catalan Government). Ethical approval was also obtained from specific speech therapy centers when required by the institution. Legal guardians were provided with written

information about the project and then provided written consent for their child to participate in the study, including video recordings of the testing and intervention sessions. They also filled out a questionnaire about their educational level (see Appendix C for the results). Additionally, all families of participating children consented to anonymized clinical data for use in scientific research.

Design

In this study, we used a non-randomized controlled trial with a between-subjects experimental design involving three different groups, which included an experimental group that received the MultiModal Narrative intervention and two control groups. Figure 1 shows a schematic representation of the study design and the intervention protocol.

Recruitment and pre-intervention assessments were completed before allocation to each group. Participants were not randomly assigned to the groups, as allocation was based on the willingness and availability of the child's speech-language therapist to implement the intervention sessions, together with families' acceptance of changing their child's usual intervention practice to the trial presented in this paper.

The MultiModal Narrative Intervention

The intensive tier of the MultiModal Narrative intervention was implemented in this study. The MultiModal Narrative is a multi-tiered narrative-based intervention for children with TD or NDD. It was co-created with 93 teachers and speech-language therapists working in the Catalan educational and clinical system, with the aim of being an evidence-based intervention program (see blinded, under review). An initial feasibility and pilot study showed that 1) professionals

found the intervention to be enjoyable, easy to implement, and to have the potential to actively engage children, and 2) children significantly improved their narrative skills after receiving the intervention sessions (see Florit-Pons et al., 2024, in press).

The intensive tier (i.e., designed for children with NDD) of the MultiModal Narrative intervention involves nine intervention sessions dedicated to training oral narrative skills through story retelling and story generation activities (see Florit-Pons et al., 2024, in press). The intervention centers on three stories about a capybara, such that each story is trained during three consecutive sessions followed by a step-by-step protocol (see Table 3 for a schematic description of the procedure). All intervention sessions start with the same four initial activities: the interventionist briefly introduces the activity and the story, then the interventionist and child watch together on a laptop the audiovisual cartoons followed by a video of a storyteller retelling and multimodally enacting the story, using representational and non-representational gestures, facial expressions and bodily movements. A central interaction-based activity follows these four initial activities where the child actively engages with the speech-language therapist. The first and second sessions' central activity involves an enacted retelling of the story using a question-and-answer sequence about the different macrostructure and perspective-taking elements: a first sequence in which the interventionist gives a first oral model of the question-answer sequence while enacting the target element using representational gestures and facial expressions; after this, the sequence is repeated so that the child and therapist can retell the story together: in this second sequence, the therapist asks the target question, the child answers, and then they both multimodally enact each element at the same time. The difference between the first two sessions is that the story is split into two: the first session focuses on the first half of the story, whereas the second session targets the second half of the story. The central activity of the third session of each story consists of a personal story generation that is related to the theme of each story, so that the child can relate what happened during the fictional stories to

their own personal experiences. First, the interventionist produces and enacts a model personal story, followed by a question-and-answer sequence; second, the interventionist asks for the child's own personal story generation and enactment. This three-session procedure is repeated for the three stories. After each intervention session, the protocol also includes the administration of two short narrative tasks to assess children's learning process throughout the intervention period. The speech-language therapist administers two short Dynamic Assessment measures (see *Session-by-session Dynamic Assessment outcome measures* subsection to see how these are implemented and coded). Children received one weekly intervention session throughout nine weeks (= nine intervention sessions). For more details on the intervention protocol, see Table 3 and Florit-Pons et al. (2024, in press).

For it to be as natural as possible for children, the intervention was implemented by the child's usual speech-language therapist in their usual intervention context. For this, all interventionists participating in this study received 6-hour training on how to implement the intervention together with specific training on multimodal enactment (e.g., different types of gestures and their value within narrative instruction). Although during the intervention sessions they had scripted instructions on when to enact the main actions and emotions and when to encourage the child to enact them, they were always advised to enact the stories as naturally as possible.

Each interventionist implemented the intervention with at least one child, although some implemented it with more than one child (M = 2; SD = 1.2; *range* = 1–4). After each intervention session, they were asked to complete a short treatment fidelity checklist indicating the duration of the session and whether they had followed the intervention protocol. The average duration of the sessions was 24.53 minutes (SD = 9.89), and on 97.32% of the occasions, they affirmed that they had followed the intervention. An additional treatment fidelity assessment was conducted by an undergraduate research assistant and either the first and third authors. Provided that each interventionist was asked to video-record a minimum of

three intervention sessions, 30% of the sessions (i.e., 48 sessions) were evaluated using a 1-7 Likert scale (see Appendix D). This external evaluation showed that interventionists followed the intervention protocol (M = 6.83, SD = 0.56), and used the set of recommended strategies (M = 6.41, SD = 1.1). Inter-rater reliability was calculated for these external fidelity assessments using Gwet's Agreement Coefficient 2 (AC2, Gwet, 2012) and showed almost perfect agreement between the coders (following the steps: AC2 = .916 (95% CI, .841 to .991), $p < .001$; using the recommended strategies: AC2 = .871 (95% CI, .814 to .928), $p < .001$).

Treatment as usual

Children in the control groups did not receive the MultiModal Narrative intervention. Children in the control NDD group continued with their usual individualized intervention sessions at the speech therapy level, whereas children in the control TD group were receiving only regular school instruction. Children's usual speech-language therapists and teachers in these control groups reported that during the target intervention weeks no oral narrative activities were conducted. Rather, they worked on other abilities: a) learning vocabulary, establishing routines, and initiating mathematical thinking at the school level (control TD), and b) phonological awareness, syntactic structure and oral conversation at the speech-therapy level (control NDD).

Outcome measures

Pre- and post-intervention outcome measures

Children's oral narrative skills were evaluated before and after the intervention in all groups using a narrative retelling task involving four stories. Two of the four stories were short (~50 *s*) wordless video cartoons about a mouse and his elephant friend from the German series "Die

Sendung mit der Maus". The third story corresponded to the first story trained during the intervention, which was also a wordless cartoon (~2.30 *min*). The fourth story was a comic-like sequence of five images from the CUBED Assessment (Petersen & Spencer, 2016). The three untrained stories (first, second, and fourth) were changed from pre-intervention to post-intervention to ensure that it was the first time the child saw that story, while the trained story was maintained post-intervention.

The procedure was as follows: the child individually watched the cartoon, and the experimenter then asked him/her to retell the story without having access to any visual material of the story. This process was repeated for the next two wordless cartoons. As for the comic-like sequence, the experimenter first told the story while referring to each picture, and then asked the child to retell it while looking at the set of pictures. Children were tested individually in a silent room at their school or speech therapy center. The task was administered and coded by a research assistant for TD children and by the first author and a certified psychologist for NDD children. Children were tested pre-intervention (around three and four weeks prior to the beginning of the intervention phase and post-intervention (around one and two weeks after the end of the intervention phase).

Each narrative retelling was coded for narrative macrostructure and narrative perspective using two different 0–6 codings (see Appendix E). The macrostructure coding was adapted from Demir et al.'s (2014) coding for evaluating the child's introduction of the different macrostructural elements (i.e., character, problem, attempt, resolution, and end) in the retelling. A score of 0 corresponded with a retelling that did not include any descriptive sequence, and a score of 6 corresponded with a retelling that included all macrostructural elements while giving details about the story. Narrative perspective-taking was coded using an adaptation of Dodd et al.'s (2011) coding, considering whether the child introduced emotional terms, causal relations of emotions, and mental terms. A score of 0 corresponded to a retelling that did not include any

emotional or mental terms, while a score of 6 corresponded to a retelling that included at least two emotions together with their cause and at least two mental terms.

Inter-rater reliability was calculated for the narrative coding of the retelling task administered pre- and post-intervention with data from 16 participants (8 TD and 8 NDD), corresponding to 32% of the data. An undergraduate third-year speech-language therapy student received 45 minutes of training on how to code participants' responses, and then annotated the data. Cohen's kappa (weighted kappa) was used to calculate inter-rater reliability. The results (*N* of responses = 128) showed almost perfect agreement with a Cohen's kappa of .86 for macrostructure skills and substantial agreement with a Cohen's kappa of .70 for perspective-taking.

Session-by-session Dynamic Assessment outcome measures

To assess children's ability to learn narrative macrostructure and perspective-taking throughout the intervention, two Dynamic Assessment measures were incorporated at the end of each intervention session: a narrative retelling task and a set of comprehension questions that incorporated graduated prompting. Graduated prompting is a scaffolded approach used to support students' learning, providing further guidance on how to respond to the task. Dynamic Assessment measures were administered and coded by the same interventionist at the end of each intervention session. Interventionists received 1-hour training before the intervention period on how to implement and code the Dynamic Assessment measures.

Figure 2 shows a visual representation of the procedure used by the practitioner to obtain the two Dynamic Assessment measures. As for the first measure, the child's narrative retelling was obtained using a set of supporting images of the trained story. The interventionist coded the child's retelling for narrative macrostructure and perspective-taking using the coding criteria described in the *Pre- and post-intervention outcome measures* subsection. Provided that during

the first intervention session for each story only the first half of the story was trained, the macrostructure score of the narrative retelling in sessions 1, 4 and 7 could only be a maximum of 4 (as a score of 5 or 6 implied that the child introduced all the elements). After the child's retelling, the interventionist asked a set of explicit graduated comprehension questions about the main macrostructure and perspective-taking elements of the trained story with different graduated support prompts (see Appendix F for a list of the comprehension questions asked for each story). She first asked an explicit open question to the child. If the child answered correctly, the response was coded as 'correct' and they moved on to the next question, but if the child answered incorrectly or did not answer, the response was coded as 'incorrect', and the interventionist would give more support prompts for the child to answer correctly, namely a two-choice question or by showing images representing the two choices. See Figure 2 for more details about the procedure.

Statistical analyses

Statistical analyses were performed using R software. First, to calculate whether there were improvements from pre- to post-intervention, four Linear Mixed-Effects (LME) models were performed using the *lme4* package (Bates et al., 2015). First, two models were run to calculate the average scores for narrative macrostructure and narrative perspective-taking for the three untrained stories. Second, two additional models were run using the scores corresponding to the trained story. The models used the narrative macrostructure score or the narrative perspective-taking score as the dependent variable. Two fixed factors were included in the models: Test (2 levels: pre-intervention and post-intervention) and Group (3 levels: Experimental NDD, Control NDD and Control TD), as well as their two-way interaction. The random-effects structure included by-Participant varying intercepts, which was determined using the *performance* package (Lüdecke et al., 2024) in R. Finally, post-hoc pairwise

comparisons were run for all significant main effects and interactions using Bonferroni correction with the *emmeans* package (Lenth, 2021), together with a measure of effect size (via Cohen's *d*).

As for the session-by-session Dynamic Assessment measures, we first ran two repeated-measures ANOVA models to examine whether there was an improvement throughout the intervention sessions. The dependent variable was the score for narrative macrostructure and narrative perspective-taking that each child received in each session. Session was included as a fixed factor, with six levels for narrative macrostructure score (Session 2, Session 3, Session 5, Session 6, Session 8 and Session 9) and nine levels for narrative perspective-taking (Sessions 1–9). Narrative macrostructure included only six levels, as the first session of each cartoon was excluded (see *Session-by-session Dynamic Assessment outcome measures* for details). Participant was included as a random factor. Post-hoc pairwise comparisons were performed using the *rstatix* package (Kassambara, 2023). Additionally, two separate LME models were used to evaluate children's ability to answer comprehension questions about the story and the support prompts they needed to answer. The first model assessing children's answers to the open questions included Score (the average number of correct and incorrect responses) as the dependent variable and Session (9 levels: Sessions 1–9), Response Type (2 levels: correct, incorrect), and Narrative Measure (2 levels: macrostructure, perspective-taking) as fixed factors. The second model assessing the support prompts needed to answer correctly included Score (the average number of support prompts) as the dependent variable and Response (4 levels: open question, two-choice question, two-choice question + image support, no correct response reached) as a fixed factor. Both models included by-Participant varying intercepts. The *emmeans* package was used to calculate post-hoc comparisons together with Cohen's *d* for the effect size.

Finally, to determine whether the support prompts that children needed could be significant predictors of later performance, a linear regression model was run. The linear regression model included the Narrative Score (i.e., the average narrative macrostructure score for untrained stories at post-test) as the dependent variable and Support Prompts (i.e., the average number of support prompts needed throughout the sessions) as a predictor.

Results

Gains in narrative macrostructure and narrative perspective-taking comparing between pre- and post-intervention

For the descriptives of all the narrative macrostructure and narrative perspective-taking measures at pre- and post-intervention, see Table 4. The model for narrative macrostructure skills for untrained stories showed a main effect of Group ($\chi^2(2) = 30.19$, $p < .001$), which indicated that, regardless of testing time, the Control TD and the Experimental group had significantly higher scores than the Control NDD group ($d = –2.46$, $p < .001$ and $d = 1.33$, $p = .016$, respectively). The interaction between Test and Group was also found to be significant ($\chi^2(2) = 10.63$, $p < .005$). Post-hoc comparisons showed first that at pre-intervention the Control TD group had significantly higher scores than the Control NDD ($d = –2.67$, $p < .001$) and the Experimental ($d = 1.91$, $p = .004$) groups. While the significant difference between the two control groups was maintained at post-intervention ($d = –2.24$, $p < .001$), we observed that the Experimental group significantly improved from pre-intervention to post-intervention ($d = 1.23$, $p < .01$) and outperformed the Control NDD group at post-intervention ($d = –1.90$, $p = .004$) (see Figure 3). As for narrative perspective-taking, results only showed a main effect of Test ($\chi^2(1) = 15.14$, $p < .001$), indicating that scores were significantly higher at pre-

intervention than at post-intervention ($d = –.77$, $p < .001$), regardless of Group. No significant effects were found for Group ($p = .468$) and for the two-way interaction ($p = .067$).

Similar findings were found for the models with the trained story, for both narrative macrostructure and perspective-taking (see Table 4). A significant two-way interaction ($\chi^2(2) = 26.72$, $p < .001$) was obtained for the macrostructure model. The interaction showed that at pre-intervention, the Control TD group had significantly higher scores than the Experimental group ($d = 2.05$, $p = .001$) and the Control NDD group ($d = –3.01$, $p < .001$), while no significant differences were found between the two groups with NDD children ($p = .592$). Additionally, results indicated that the Experimental group significantly improved from pre-intervention to post-intervention ($d = 1.63$, $p < .001$), and that at post-test, scores were significantly higher than those of the Control NDD group ($d = –1.68$, $p = .047$). As for narrative perspective-taking skills for the trained story, results showed a significant two-way interaction between Test and Condition ($\chi^2(2) = 6.80$, $p = .033$), but post-hoc comparisons were not statistically significant, showing no differences between the groups at the two testing times.

Gains in narrative macrostructure and narrative perspective-taking using Dynamic Assessment measures

As mentioned before, two basic Dynamic Assessment measures were collected at the end of each intervention session, namely a narrative retelling and a set of comprehension questions with graduated prompting. Concerning the session-by-session analysis of the Dynamic Assessment narrative retellings, the ANOVA model evaluating narrative macrostructure showed a main effect of Session ($F(5, 70) = 2.53$, $p = .036$), showing that there was a significant improvement from Session 2 to Session 3 ($t(13) = –4.20$, $p = .015$) and from Session 2 to Session 5 ($t(14) = –4.00$, $p = .02$). No significant differences were found for the rest of the

sessions, suggesting that the improvement was maintained throughout the remaining intervention sessions. The model for narrative perspective-taking also showed a main effect of Session ($F(8, 114) = 3.68$, $p < .001$), which reported significant improvements from Session 1 to Session 6 ($t(14) = –4.58$, $p = .015$), from Session 1 to Session 8 ($t(15) = –4.37$, $p = .02$), and from Session 1 to Session 9 ($t(13) = –4.16$, $p = .04$), showing that it took more time for children to acquire this skill. See Table 5 for the descriptive statistics of all sessions and Figure 4 for a representation of these improvements.

Second, regarding the analysis of the comprehension prompting questions that defined the support needed by the children, the first LME model evaluating their ability to answer open questions correctly or incorrectly showed a significant main effect of Response Type ($\chi^2(1) = 30.92$, $p < .001$), indicating that overall there were significantly more correct answers than incorrect answers ($d = .58$, $p < .001$). No main effects were found for Session ($p = 1$) or Narrative Measure (i.e., macrostructure, perspective-taking) ($p = .998$). The second LME evaluating the support prompts needed to answer correctly showed a main effect of Response ($\chi^2(3) = 93.02$, $p < .001$) which showed that most children could reach a correct response after the hierarchy of support prompts, such as the open question ($d = –3.21$, $p < .001$) or the two-choice question ($d = –1.43$, $p = .001$). It also showed that children could significantly answer more often at the first prompt (i.e., open question) than at the other graduated prompts, such as two-choice question ($d = 1.78$, $p < .001$) or two-choice question + image support ($d = 2.59$, $p < .001$), independently of whether it was a macrostructure or perspective-taking question. All in all, these findings suggest that although children needed different support prompts, overall they could answer all questions correctly. See Table 6 for the percentages of correct/incorrect responses given and the support prompts needed.

Finally, the linear regression model to assess whether the average number of graduated support prompts could predict narrative gains at post-test showed that the model accounted for 52.12%

of the variance (R^2 = .52, F(1,14) = 17.33, p < .001), suggesting that the average number of support prompts children needed to answer correctly was a significant predictor of children's narrative macrostructure score on the untrained stories at the post-intervention assessment (β = –.89, p < .001) (see Figure 5).

Discussion

The current study evaluated the efficacy of the MultiModal Narrative intervention on a group of children with NDD. In particular, the intensive version of the intervention was used, which crucially incorporates a set of validated instructional strategies including a strong multimodal component, and it is naturalistic (implemented by the child's usual speech-language therapist). Specifically, we aimed to assess its effectiveness by comparing children's oral narrative performance before and after the intervention through narrative retellings and during the intervention through Dynamic Assessment measures. As a result of the intervention, we found, first, that children who received the intervention significantly improved their narrative macrostructure skills (for both the trained and the untrained stories), but not their narrative perspective-taking skills. More precisely, when looking at the session-by-session learning process, we observed that narrative macrostructure was learned during the first sessions of the intervention, while it took longer for children to learn narrative perspective-taking skills. Our results showed that even though children could answer questions about the story, around 40% of the time they needed extra support prompts, such as open questions, two-choice questions, or image support to answer the questions correctly. Crucially, the amount of support that children needed was a significant predictor of their narrative macrostructure performance in the post-intervention assessment.

The findings of this study contribute to the existing narrative intervention literature by providing evidence of the effectiveness of individualized multimodal-based narrative instruction on macrostructure skills. We believe that a variety of factors have contributed to the effectiveness of the MultiModal Narrative intervention. First, the MultiModal Narrative intervention complies with the characteristics of effective interventions described in the review by Favot and colleagues (2021), which are also in line with some of the principles of narrative intervention described by Spencer and Petersen (2020), such as the use of visual materials like story icons, modeling from the speech-language therapist, and retelling the story during each intervention session (for more details on the MultiModal Narrative intervention, see Florit-Pons et al., 2024, in press). Second, the MultiModal Narrative intervention not only incorporates these characteristics but also systematically includes multimodal strategies, such as the use of story enactment to visually represent the main actions and emotions of the story. A great amount of research highlights the causal role of multimodality in children's language skills (for reviews see e.g., Rohlfing, 2019; Vilà-Giménez & Prieto, 2021), supported by the multimodal enrichment paradigm (Mathias & von Kriegstein, 2023). Finally, the MultiModal Narrative intervention was co-created with professionals to ensure that it integrated professionals' needs and real-life intervention practice (see blinded, under review), and, its feasibility and preliminary effectiveness were evaluated in a feasibility pilot study (see Florit-Pons et al., 2024, in press). We believe that these two complementary actions might have helped tailor the intervention to the Catalan context, thus making it more effective. Ensuring that interventions are developed considering the implementation context makes them sensitive and adapted to the professionals' needs. Therefore, this emphasizes the need to test the effectiveness of narrative interventions in different contexts, as although the core intervention principles might be transferable to different languages, the contextual factors might vary across different professional settings in different countries and regions.

Nevertheless, the effectiveness of the MultiModal Narrative intervention in boosting narrative perspective-taking skills is not clearcut. While we observed that children were slowly learning this skill throughout the intervention using Dynamic Assessment measures, this improvement was not observed during the post-intervention assessment. First, it is important to consider that children with NDD usually have difficulties comprehending and expressing their thoughts and emotions (see Fujiki & Brinton, 2017 for a review). In addition, although children can identify and verbalize basic emotions by the ages of 3–4 in emotion identification tasks (for a review, see Widen & Nelson, 2022), we believe that identifying and expressing them in a narrative discourse might be more complex. Building a narrative discourse requires multiple cognitive and linguistic processes to function simultaneously, such as memorizing and organizing both the story events and characters' emotions, thoughts and perspectives, finding the appropriate words to describe these events and mental terms, putting them into the correct syntactic structure, and then expressing them cohesively within the narrative discourse while considering multiple perspectives (e.g., García-Pérez et al., 2008). In fact, the difficulty of children with NDD to identify and express the character's emotions and perspectives within a narrative discourse has already been documented in previous research (for reviews, see e.g., Baixauli et al., 2016; Norbury et al., 2014; Winters et al., 2022). Therefore, we believe that children at these early ages (particularly children with NDD) might struggle to deal with all these processes at the same time, and for this, they avoid talking about emotions and thoughts, which are less salient components within the narrative. We believe that significant effects of the intervention on perspective-taking skills would have been found if children were older or had received more specific instruction on narrative perspective-taking (such as the one by Dodd et al., 2011). As suggested by the results of our Dynamic Assessment measures, it seemed to take longer for children to acquire narrative perspective-taking skills. In this regard, a longer intervention

would have helped ensure a boosting effect on perspective-taking skills at the end of the intervention.

The current study has shown that using multiple Dynamic Assessment measures throughout the intervention helps to understand children's learning abilities. Specifically, we observed that children with NDD needed different support prompts, such as two-choice questions or visuals, to answer correctly. Our findings have shown that Dynamic Assessment measures, specifically the assessment of the support prompts that each child needs, predict children's post-intervention outcomes (in line with Camilleri & Botting, 2013; Hasson et al., 2012; Olswang & Bain, 1996), indicating that it is a relevant and valid measure for intervention practice, especially narrative intervention. Overall, the inclusion of these measures in narrative interventions is useful to test the learning capability of each child during the intervention and to adapt the intervention targets and procedure according to the child's needs. Continuously assessing the amount of support prompts children need to learn a certain linguistic skill, such as narrative macrostructure or perspective-taking, will serve to predict future outcomes and adapt the intervention accordingly.

It is necessary to acknowledge some relevant limitations of the present study. While our study provides valuable insights on the preliminary efficacy of the MultiModal Narrative intervention, the sample size of this study was small compared to some existing large-scale randomized controlled trials. The small sample size in this study can be attributed to several factors. First, it was difficult to recruit a larger and more homogeneous sample, mostly because the experimental group required double participation (i.e., the child and the therapist implementing the intervention). Although the initial age target was children aged four to six (which is when children's ability to narrate a story starts establishing), the double participation requirement made us have to widen the age range in order to be able to recruit enough participants. Despite this, it was still difficult to recruit a sample that contained a balanced

number of children with autism and children with Developmental Language Disorder. Second, there was also a 9.4% dropout rate (see Appendix A) for children who did not meet the inclusion criteria or abandoned the study during the intervention period for various reasons, such as recurrent illness or difficulties in arranging the post-intervention session with the family. Overall, although the sample size was determined by an *a priori* power analysis, we acknowledge that the sample size is small, thus restricting the generalizability of the findings to a larger population and limiting the ability to draw broad conclusions. It should also be acknowledged that the power analysis was conducted specifically for the models assessing narrative gains between pre- and post-intervention, but not for the models assessing the Dynamic Assessment measures, which were only administered to the experimental group. Another limitation of this study is that inter-rater reliability could not be calculated for the Dynamic Assessment measures, provided that not all of the intervention sessions were video recorded and that some interventionists stopped the recordings right before administering these measures. A final limitation of this study is that no background information (e.g., education, years of experience) was collected from the speech-language therapists implementing the intervention.

Finally, our study has left some questions for future research. Although one of the core aspects of the intervention is the multimodal component, we did not directly assess the individual contribution that it might have to the intervention, because of difficulties with participant recruitment. We believe that it would be useful that future studies evaluate the relative contribution of the multimodal component within a narrative intervention to the improvements in the narrative outcome measures. In addition, we believe that future studies should analyze children's narrative perspective-taking skills more deeply, for instance by distinguishing between emotional and mental terms, as well as the causal relations behind those terms. This

more fine-grained analysis would serve to identify which terms children used more frequently and in which area was there an improvement.

The current study has positive implications for speech-language pathology research and practice. First, the study has shown that adopting evidence-based interventions co-created with professionals helps ensure that the intervention is suitable for both children receiving the intervention and professionals implementing it. Second, the MultiModal Narrative intervention protocol includes different educational strategies for improving oral discourse, such as verbal modeling, feedback, use of audiovisual support, and story enactment. The systematic incorporation of these strategies makes interventionists more aware of the aim and effect of each strategy and allows them to use them in other intervention contexts. In line with this, a direct and practical outcome of this study is the materialization of the MMN, readily accessible to professionals who lack scientifically validated narrative intervention materials in the Catalan language (blinded, under review). Finally, because speech-language therapists do not usually implement pre- and post-intervention assessments, as they are time-consuming, having Dynamic Assessment measures can be especially helpful. The fact that the intensive implementation of the MultiModal Narrative intervention incorporates Dynamic Assessment can help them quickly assess whether children are following the intervention and learning, or whether they need more support and the pace of intervention needs to be adapted.

Conclusion

This study reports on the effectiveness of a 9-session individualized multimodal narrative intervention for boosting NDD children's narrative skills and emphasizes the importance of using dynamic assessment measures throughout the intervention process to evaluate children's continuous learning ability.

References

Acosta-Rodríguez, V. M., Ramírez Santana, G. M., Cruz, A. D., & Del Valle Hernández, Y. N. (2022). Trastorno Específico del Lenguaje y Trastorno del Espectro Autista: Similitudes y diferencias en el discurso narrativo [Specific Language Impairment and Autism Spectrum Disorder: similarities and differences in narrative discourse]. *Revista de Investigación en Logopedia*, *12*(1), Article e76069. https://doi.org/10.5209/rlog.76069

Andrés-Roqueta, C., Flores-Builes, R., & Igualada, A. (2024). Validation of PleaseApp: a digital tool for the assessment of receptive pragmatic abilities in children with neurodevelopmental disorders. *Frontiers in Psychiatry*, *15*, Article 1329022. https://doi.org/10.3389/fpsyt.2024.1329022

Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry*, *63*(4), 397–417. https://doi.org/10.1111/jcpp.13481

Babayiğit, S., Roulstone, S. & Wren, Y. (2021). Linguistic comprehension and narrative skills predict reading ability: A 9-year longitudinal study. *British Journal of Educational Psychology*, *91*, 148–168. https://doi.org/10.1111/bjep.12353

Baixauli, I., Colomer, C., Roselló, B., & Miranda, A. (2016). Narratives of children with high-functioning autism spectrum disorder: A meta-analysis. *Research in Developmental Disabilities*, *59*, 234–254. https://doi.org/10.1016/j.ridd.2016.09.007

Bamford, C. K., Masso, S., Baker, E., & Ballard, K. J. (2022). Dynamic assessment for children with communication disorders: A systematic scoping review and framework. *American Journal of Speech-Language Pathology*, *31*(4), 1878–1893. https://doi.org/10.1044/2022_ajslp-21-00349

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bernstein, K. A., van Huisstede, L., Marley, S. C., Gao, Y., Pierce-Rivera, M., Ippolito, E., Restrepo, M. A., Millinger, J., Brantley, K., & Gantwerker, J. (2024). Gesture like a kitten and you won't forget your tale: Drama-based, embodied story time supports preschoolers' narrative skills. *Early Childhood Research Quarterly*, *66*, 178–190. https://doi.org/10.1016/j.ecresq.2023.10.004

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLOS ONE*, *11*(7), Article e0158753. https://doi.org/10.1371/journal.pone.0158753

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 consortium (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Camilleri, B., & Botting, N. (2013). Beyond static assessment of children's receptive vocabulary: The dynamic assessment of word learning (DAWL). *International Journal of Language & Communication Disorders*, *48*(5), 565–581. https://doi.org/10.1111/1460-6984.12033

Clark, A. G., & Dockweiler, K. A. (2020). *Multi-Tiered Systems of Support in Elementary Schools: The Definitive Guide to Effective Implementation and Quality Control*. Routledge.

Dargue, N., Sweller, N., & Carter, M. (2021). Short report: Learning through iconic gesture in autism spectrum disorder. *Research in Developmental Disabilities*, *115*, Article 104000. https://doi.org/10.1016/j.ridd.2021.104000

Delgado-Cruz, A., Acosta-Rodríguez, V. M., & Ramírez-Santana, G. M. (2022). Intervention in the coherence of narrative discourse in students with Developmental Language Disorder and with Typical Development (*Intervención en la coherencia del discurso narrativo de alumnado con Trastorno del Desarrollo del Lenguaje y con Desarrollo Típico*). *Journal for the Study of Education and Development*, *45*(4), 774–803. https://doi.org/10.1080/02103702.2022.2096279

Delgado-Cruz, A., Acosta-Rodríguez, V. M., & Ramírez-Santana, G. M. (2024). Enseñanza de esquemas narrativos en alumnado de Educación Infantil: trastorno del desarrollo del lenguaje y desarrollo típico. *Estudios sobre Educación*, *46*, 99–120. https://doi.org/10.15581/004.46.005

Demir, Ö. E., Fisher, J. A., Goldin-Meadow, S., & Levine, S. C. (2014). Narrative processing in typically developing children and children with early unilateral brain injury: Seeing gesture matters. *Developmental Psychology*, *50*(3), 815–828. https://doi.org/10.1037/a0034322

Dickinson, D. K., & McCabe, A. (2001). Bringing it all together: The multiple origins, skills, and environmental supports of early literacy. *Learning Disabilities Research & Practice*, *16*(4), 186–202.

Dodd, J. L., Ocampo, A., & Kennedy, K. S. (2011). Perspective taking through narratives: An intervention for students with ASD. *Communication Disorders Quarterly*, *33*(1), 23–33. https://doi.org/10.1177/1525740110395014

Donolato, E., Toffalini, E., Rogde, K., Nordahl-Hansen, A., Lervåg, A., Norbury, C., & Melby-Lervåg, M. (2023). Oral language interventions can improve language outcomes in children with neurodevelopmental disorders: A systematic review and meta-analysis. Campbell systematic reviews, *19*, Article e1368. https://doi.org/10.1002/cl2.1368

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Favot, K., Carter, M., & Stephenson, J. (2021). The effects of oral narrative intervention on the narratives of children with language disorder: A systematic literature review. *Journal of Developmental and Physical Disabilities*, *33*(4), 489–536. https://doi.org/10.1007/s10882-020-09763-9

Fiestas, C. E., & Peña, E. D. (2018). The dynamic assessment of narratives: A bilingual study. *Journal of Cognitive Education and Psychology*, *17*(1), 97–111. https://doi.org/10.1891/1945-8959.17.1.97

Florit-Pons, J., Igualada, A., & Prieto, P. (2024, in press). Evaluating the feasibility and preliminary effectiveness of a multi-tiered multimodal narrative intervention program for preschool children. *Language, Speech, and Hearing Services in Schools*.

blinded. (under review). Co-creation of a narrative intervention program for speech–language pathology and educational settings.

Frey, N., & Lüke, C. (2023). Iconic gestures support novel word learning in multilingual students with SLCN in classrooms. *Child Language Teaching and Therapy*, *39*(2), 119–134. https://doi.org/10.1177/02656590231166919

Fujiki, M., Brinton, B. (2017) Social communication intervention for children with language impairment. In: R. McCauley, M. E. Fey, & R. Gillam (Eds.), *Treatment of language disorders* (pp. 421–449). Brookes Publishing.

García-Pérez, R. M., Hobson, R. P., & Lee, A. (2008). Narrative role-taking in autism. *Journal of Autism and Developmental Disorders*, *38*, 156–168. https://doi.org/10.1007/s10803-007-0379-z

Gwet, K. L. (2012). *Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among multiple raters*. Advanced Analytics Press.

Hasson, N., Dodd, B., & Botting, N. (2012). Dynamic Assessment of Sentence Structure (DASS): Design and evaluation of a novel procedure for the assessment of syntax in children with language impairments. *International Journal of Language & Communication Disorders*, *47*(3), 285–299. https://doi.org/10.1111/j.1460-6984.2011.00108.x

Hessling, A., & Schuele, C. M. (2020). Individualized Narrative Intervention for School-Age Children With Specific Language Impairment. *Language, Speech, and Hearing Services in Schools*, *51*(3), 687–705. https://doi.org/10.1044/2019_LSHSS-19-00082

Ireland, M. C., & Hall-Mills, S. (2024). Empowering Speech-Language Pathologists: Strategies for Effective Individualized Education Program Navigation and Inclusive Practice in Schools. *Language, Speech, and Hearing Services in Schools*, *55*(2), 1–6. https://doi.org/10.1044/2024_LSHSS-24-00026

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (Eds.). (2016). *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed.). Springer Science + Business Media. https://doi.org/10.1007/978-1-4899-7568-3

Kassambara A (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests.* R package version 0.7.2. https://rpkgs.datanovia.com/rstatix/

Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test*. Pearson.

Kuhl, P. K. (2014). Early Language Learning and the Social Brain. *Cold Spring Harbor Symposia on Quantitative Biology*, *79*, 211–220. https://doi.org/10.1101/sqb.2014.79.024802

Lenth, R. V. (2021). *emmeans: Estimated marginal means, aka least-squares means* (R Package Version 1.5.5-1). https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), Article 3139. https://doi.org/10.21105/joss.03139

Lytle, S. R., & Kuhl, P. K. (2017). Social Interaction and Language Acquisition. In E. M. Fernández & H. Smith Cairns (Eds.) *The Handbook of Psycholinguistics* (pp. 615–634). John Wiley & Sons. https://doi.org/10.1002/9781118829516.ch27

Norbury, C. F., & Bishop, D. V. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287–313. https://doi.org/10.1080/1368203100001081 33

Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language, 41*(3), 485–510. https://doi.org/10.1017/S030500091300007X

Mathias, B., & von Kriegstein, K. (2023). Enriched learning: Behavior, brain, and computation. *Trends in Cognitive Sciences*, *27*(1), 81–97. https://doi.org/10.1016/j.tics.2022.10.007

McGregor K. K. (2020). How We Fail Children With Developmental Language Disorder. *Language, Speech, and Hearing Services in Schools*, *51*(4), 981–992. https://doi.org/10.1044/2020_LSHSS-20-00003

Nicolopoulou, A., Cortina, K. S., Ilgaz, H., Cates, C. B., & de Sá, A. B. (2015). Using a narrative- and play-based activity to promote low-income preschoolers' oral language, emergent literacy, and social competence. *Early Childhood Research Quarterly*, *31*, 147–162. https://doi.org/10.1016/j.ecresq.2015.01.006

Olswang, L. B., & Bain, B. A. (1996). Assessment Information for Predicting Upcoming Change in Language Production. *Journal of Speech, Language, and Hearing Research*, *39*(2), 414–423. https://doi.org/10.1044/jshr.3902.414

Orellana, C. I., Wada, R., & Gillam, R. B. (2019). The Use of Dynamic Assessment for the Diagnosis of Language Disorders in Bilingual Children: A Meta-Analysis. *American journal of speech-language pathology*, *28*(3), 1298–1317. https://doi.org/10.1044/2019_AJSLP-18-0202

Peña, E. D., Gillam, R. B., & Bedore, L. M. (2014). Dynamic assessment of narrative ability in English accurately identifies language impairment in English language learners. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2208–2220. https://doi.org/10.1044/2014_JSLHR-L-13-0151

Peña, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: an experimental investigation of classification accuracy. *Journal of Speech, Language, and Hearing Research*, *49*(5), 1037–1057. https://doi.org/10.1044/1092-4388(2006/074)

Petersen, D. B., Chanthongthip, H., Ukrainetz, T. A., Spencer, T. D., & Steeve, R. W. (2017). Dynamic Assessment of Narratives: Efficient, Accurate Identification of Language Impairment in Bilingual Students. *Journal of Speech, Language, and Hearing Research*, *60*(4), 983–998. https://doi.org/10.1044/2016_JSLHR-L-15-0426

Petersen, D. B., & Spencer, T. D. (2016). *CUBED*. Language Dynamics Group.

Pico, D. L., Hessling Prahl, A., Biel, C. H., Peterson, A. K., Biel, E. J., Woods, C., & Contesse, V. A. (2021). Interventions Designed to Improve Narrative Language in School-Age Children: A Systematic Review With Meta-Analyses. *Language, Speech, and Hearing Services in Schools*, *52*(4), 1109–1126. https://doi.org/10.1044/2021_LSHSS-20-00160

Rohlfing, K. (2019). Learning language from the use of gestures. In J. S. Horst & J. V. Torkildsen (Eds.), *International Handbook of Language Acquisition* (pp. 213–233). Routledge.

Spencer, T. D., & Petersen, D. B. (2018). Bridging Oral and Written Language: An Oral Narrative Language Intervention Study With Writing Outcomes. *Language, Speech, and Hearing Services in Schools*, *49*(3), 569–581. https://doi.org/10.1044/2018_LSHSS-17-0030

Spencer, T. D., & Petersen, D. B. (2020). Narrative Intervention: Principles to Practice. *Language, Speech, and Hearing Services in Schools*, *51*(4), 1081–1096. https://doi.org/10.1044/2020_LSHSS-20-00015

Vilà-Giménez, I., & Prieto, P. (2021). The Value of Non-Referential Gestures: A Systematic Review of Their Cognitive and Linguistic Effects in Children's Language Development. *Children*, *8*(2), Article 148. https://doi.org/10.3390/children8020148

Whalon, K., Henning, B., Jackson, E., & Intepe-Tingir, S. (2019). Effects of an adapted story grammar intervention on the listening comprehension of children with autism. *Research in developmental disabilities*, *95*, Article 103507. https://doi.org/10.1016/j.ridd.2019.103507

Widen, S. C., & Nelson, N. L. (2022). Differentiation and Language Acquisition in Children's Understanding of Emotion. In D. Dukes, A. C. Samson & E. A. Walle (Eds.), *The Oxford Handbook of Emotional Development* (pp. 174–187). Oxford Academic.

Wiig, E. H., Secord, W. A. & Semel, E. (2009). *Clinical Evaluation of Language Fundamentals–Preschool (CELF-2)*. Pearson.

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals–fifth edition (CELF-5)*. Pearson.

Winters, K. L., Jasso, J., Pustejovsky, J. E., & Byrd, C. T. (2022). Investigating Narrative Performance in Children With Developmental Language Disorder: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *65*(10), 3908–3929. https://doi.org/10.1044/2022_JSLHR-22-00017

Tables and figures

Table 1

Participants' characteristics

Variable	Control TD	Control NDD	Experimental NDD
N of participants (*N* of females and males)	17 (8 F, 9 M)	17 (5 F, 12 M)	16 (6 F, 10 M)
N of participants with a diagnosis (*N* of F and M)	–	9 (2 F, 7 M)	9 (1 F, 8 M)
N of participants with reported risk (*N* of F and M)	–	8 (3 F, 5 M)	7 (5 F, 2 M)
Age: Mean (SD)	5.45 (0.24)	4.75 (0.78)	5.20 (0.96)
Age: Range	5.08–5.75	3.92–7.08	3.5–7.33
TSI[1]: Mean (SD)	102 (0)	95.06 (9.24)	102.43 (12.04)
TSI: Range	102	87.7–110.8	70.5–120.2
CELF: Mean (SD)	93.06 (12.0)	82.94 (15.33)	81.38 (13.67)
CELF: Range	76–115	58–106	58–105
PleaseApp: Mean (SD)	29.65 (6.36)	19.71 (4.90)	24.00 (6.99)

[1] The territorial socioeconomic index (TSI) is a value calculated by the Statistical Institute of Catalonia that summarizes the socioeconomic profile of a population in a certain area. The average IST value for the region of Catalonia is 100, with values ranging from 36.5 to 137.1. For this study, we used this value calculated at the municipality level (considering different neighborhoods within big cities) where the school or speech-therapy center that the children were attending was located. For more information see https://www.idescat.cat/pub/?id=ist&lang=en

PleaseApp: Range	19–42	10–30	13–38
K-BIT: Mean (SD)	103.06 (13.44)	101.65 (13.26)	101.56 (13.70)
K-BIT: Range	83–123	74–123	82–127

Table 2

Summary of the t-test results and Cohen's d *by Group*

Variable	**Experimental NDD vs. Control TD**	**Experimental NDD vs. Control NDD**	**Control TD vs. Control NDD**
Age	$t(16.83) = -1.00$, $p = .330$, $d = 0.36$	$t(28.90) = 1.50$, $p = .144$, $d = 0.53$	$t(19.15) = 3.57$, $p = .002$, $d = 1.23$
TSI	$t(15.00) = 0.143$, $p = .888$, $d = 0.05$	$t(28.13) = 1.96$, $p = .059$, $d = 0.69$	$t(16.00) = 3.01$, $p = .007$, $d = 1.06$
CELF	$t(29.90) = -2.60$, $p = .014$, $d = -0.91$	$t(30.92) = -0.31$, $p = .759$, $d = -0.11$	$t(30.25) = 2.14$, $p = .040$, $d = 0.74$
PleaseApp	$t(30.27) = -2.42$, $p = .022$, $d = -0.85$	$t(26.72) = 2.03$, $p = .052$, $d = 0.72$	$t(30.03) = 5.11$, $p < .001$, $d = 1.75$
K-BIT	$t(30.79) = -0.32$, $p = .754$, $d = -0.11$	$t(30.72) = -0.02$, $p = .986$, $d = -0.01$	$t(31.99) = 0.31$, $p = .760$, $d = 0.11$

Figure 1

Study design and schematic representation of the study design and intervention protocol

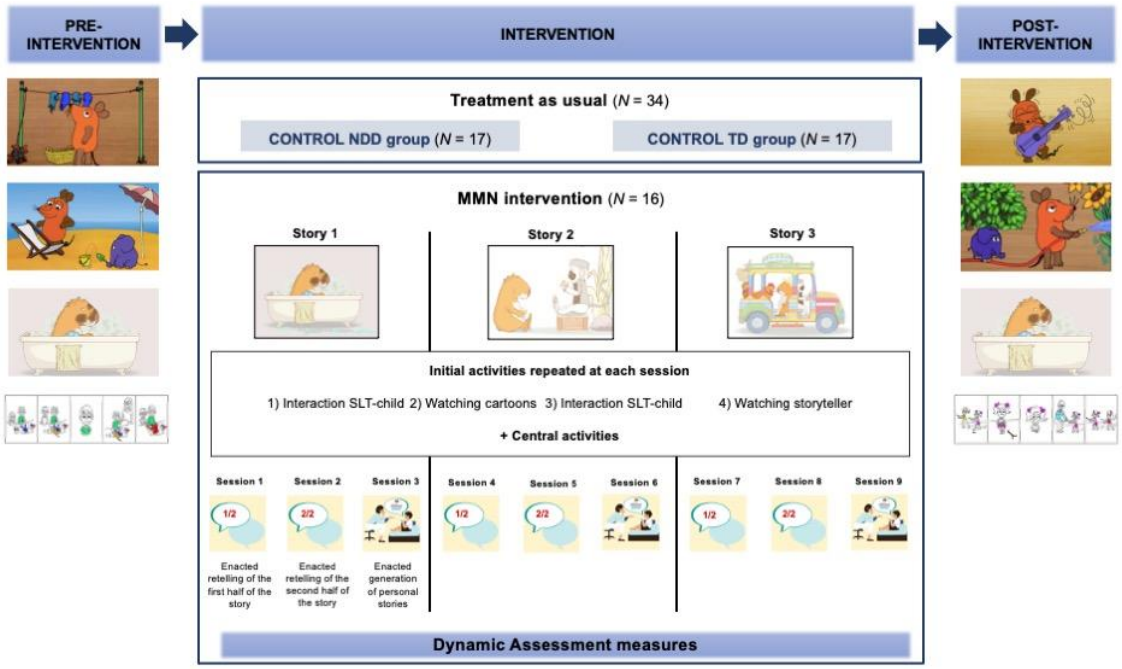


Table 3

MultiModal Narrative intervention procedure

Activity	Step-by-step description of the procedure
Initial activities repeated at each session	1. One-to-one therapist-child interaction presenting the session. 2. Watching wordless cartoons. 3. One-to-one therapist-child interaction briefly asking about the story plot. 4. Watching a video of a storyteller retelling and multimodally enacting the story.

Sessions 1-2. Enacted retelling of the story	Use of question-and-answer (Q&A) sequences targeting the main story elements (i.e., protagonist, initial emotion, problem, emotional reaction after the problem, attempt to solve the problem, emotional reaction after the attempt, solution, emotion after the solution and end of the story), accompanied by visual supports (story icons and short animated video representing each story element). 1. Q&A sequence 1: Therapist asks and immediately answers all the questions herself. When answering, she enacts the main action or emotion 2. Q&A sequence 2: 2.1. Therapist asks a question and the child answers. 2.2. Therapist provides positive feedback to the child. 2.3. Therapist enacts the story element corresponding to that question. 2.4. Therapist encourages the child to also enact it. 2.5. Process repeated for all questions.
Session 3. Enacted generation of personal story	1. Therapist generates and enacts her own story. 2. Q&A sequence 1: Therapist asks a question about her own story and the child answers. 3. The child generates and enacts her own story. 4. Q&A sequence 1: Therapist asks a question about the child's personal story and the child answers.

Figure 2

Visual representation of the procedure used to obtain the Dynamic Assessment measures

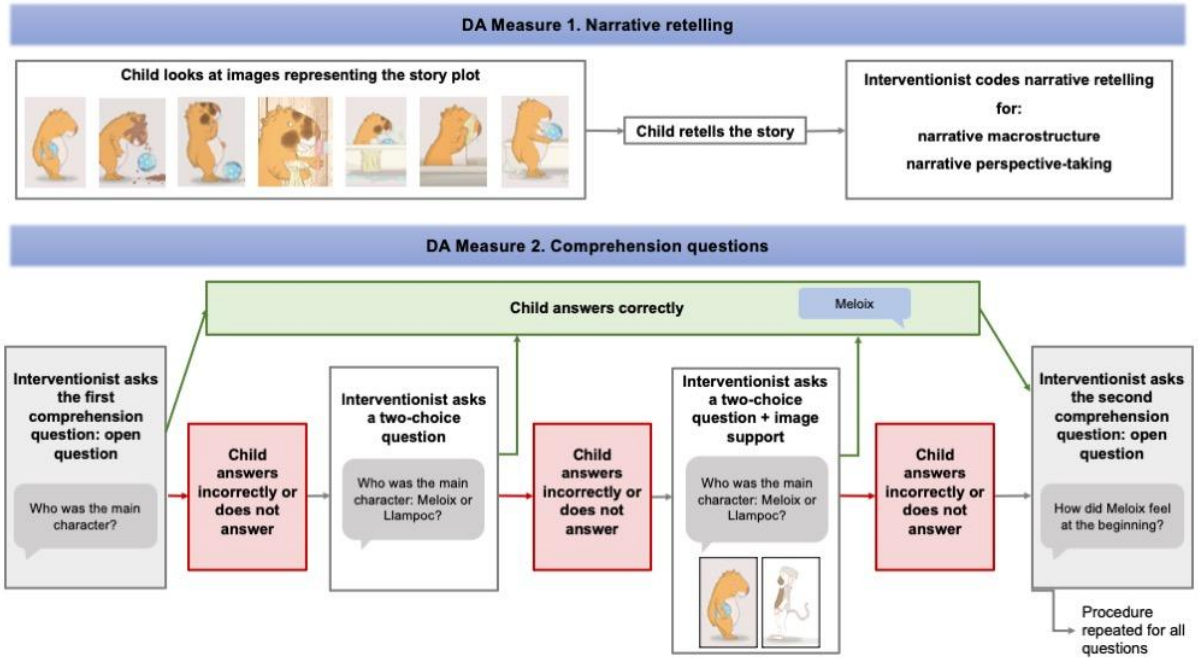


Table 4

*Descriptives (*M, SD *and* Range*) for pre- and post-intervention assessments*

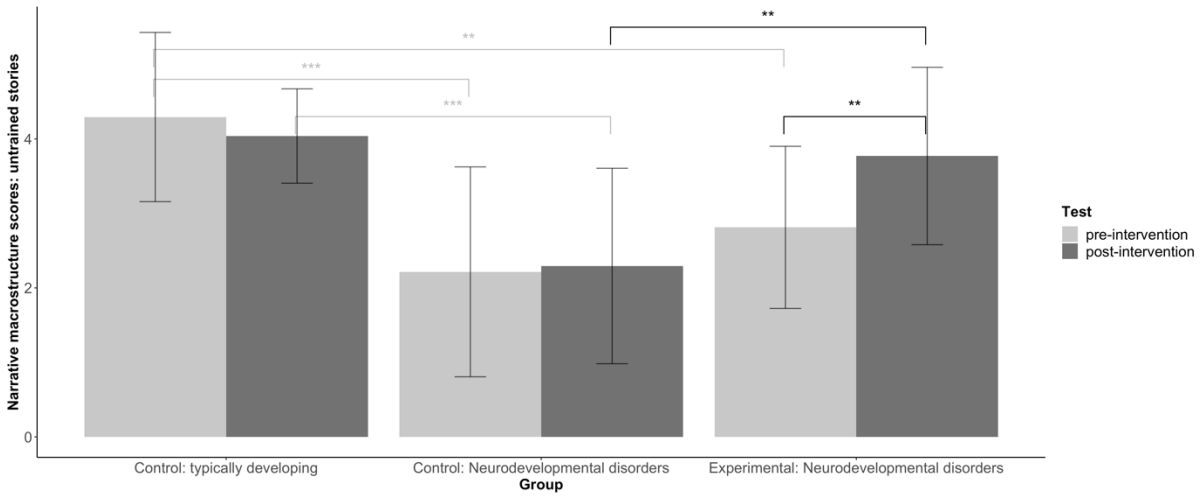
			Control TD		Control NDD		Experimental NDD	
Variable	**Story**		**PRE**	**POST**	**PRE**	**POST**	**PRE**	**POST**
Narrative macrostructure	**Trained story**	*M*	4.35	3.71	1.82	2.76	2.63	4.00
		SD	0.70	0.92	1.78	1.35	1.02	1.32
		Range	3–5	2–6	0–5	0–5	1–4	1–6
	Untrained	*M*	4.29	4.04	2.22	2.29	2.81	3.77

	stories	*SD*	1.14	0.63	1.41	1.31	1.09	1.19
		Range	0.67–5.67	3–5.33	0–4.67	0–4.33	1–4	1.67–5.67
Narrative perspective-taking	**Trained story**	*M*	0	0	0.24	0.06	0	0.31
		SD	0	0	0.75	0.24	0	0.60
		Range	–	–	0–3	0–1	–	0–2
	Untrained stories	*M*	0.39	0.06	0.29	0.02	0.26	0.23
		SD	0.40	0.13	0.44	0.08	0.29	0.23
		Range	0–1.33	0–0.33	0–1.67	0–0.33	0–0.67	0–0.67

Note. See Appendix G for this table separating participants with autism and Developmental Language Disorder

Figure 3

Mean narrative macrostructure scores for the untrained stories broken down by Test (pre-intervention and post-intervention) and Group (Control TD, Control NDD and Experimental NDD).



Note. Asterisks represent significant differences: * stands for $p \leq .05$; ** stands for $p \leq .01$; and *** stands for $p \leq .001$.

Table 5

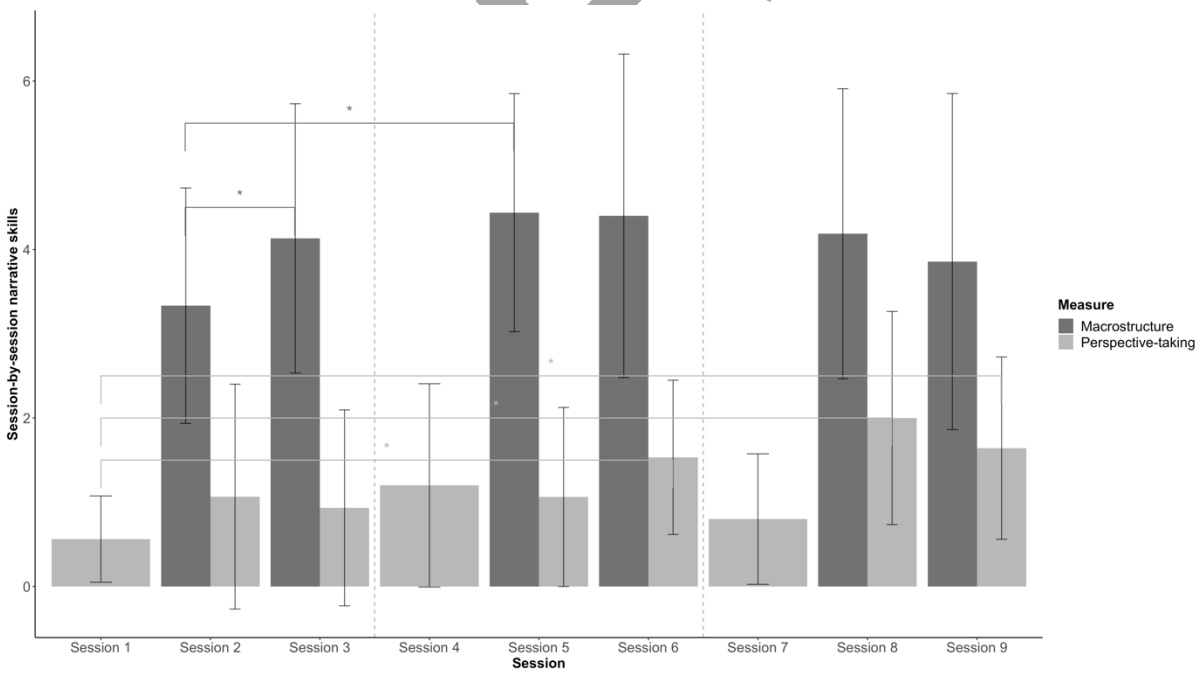
*Descriptives (*M, SD *and* Range*) for the session-by-session narrative retelling*

		Session								
Variable		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Narrative macrostructu	*M*	–	3.33	4.13	–	4.44	4.40	–	4.19	3.86
	SD	–	1.40	1.60	–	1.41	1.92	–	1.72	1.99

re	*Rang e*	–	0–5	0–6	–	2–6	1–6	–	1–6	1–6
Narrative perspective-taking	*M*	0.56	1.07	0.93	1.20	1.06	1.53	0.80	2.00	1.64
	SD	0.51	1.33	1.16	1.21	1.06	0.92	0.77	1.26	1.08
	Rang e	0–1	0–4	0–4	0–4	0–4	0–3	0–2	0–4	0–3

Figure 4

Mean narrative macrostructure and perspective-taking scores for the session-by-session narrative retelling task broken down by Narrative Measure (macrostructure and perspective-taking) and Session (Session 1–9).



Note. Asterisks represent significant differences: * stands for $p \leq .05$; ** stands for $p \leq .01$; and *** stands for $p \leq .001$.

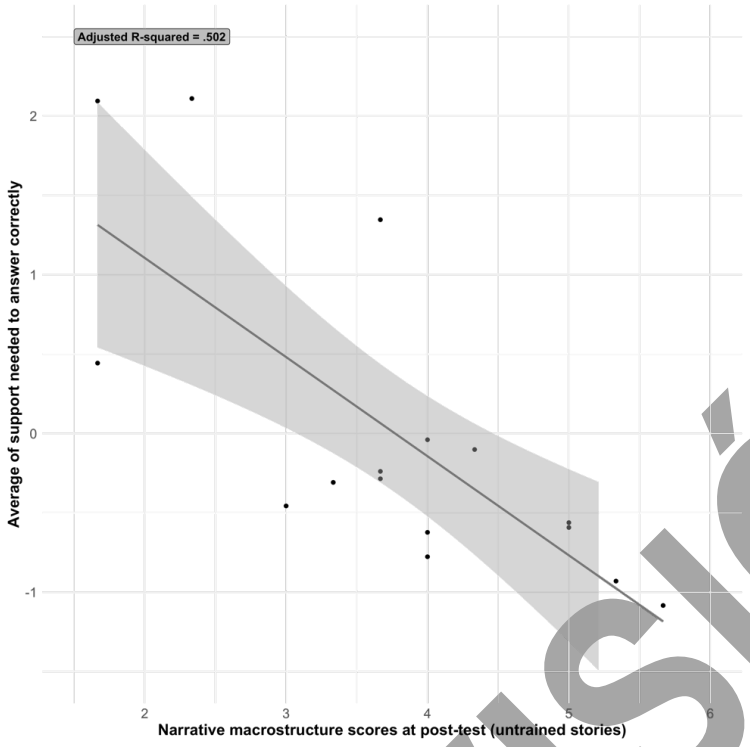
Table 6

Percentage of responses according to each Response Type during the two phases of the graduated prompting comprehension questions

Responses	**%**
First step	
correct responses	60.06%
incorrect responses	39.94%
Second step	
open question	60.68%
two-choice question	27.12%
two-choice question + image support	11.97%
no correct response reached	0.23%

Figure 5

The relation between narrative macrostructure skills at post-intervention and the average number of support prompts needed throughout the intervention, as estimated by multiple regression analysis.

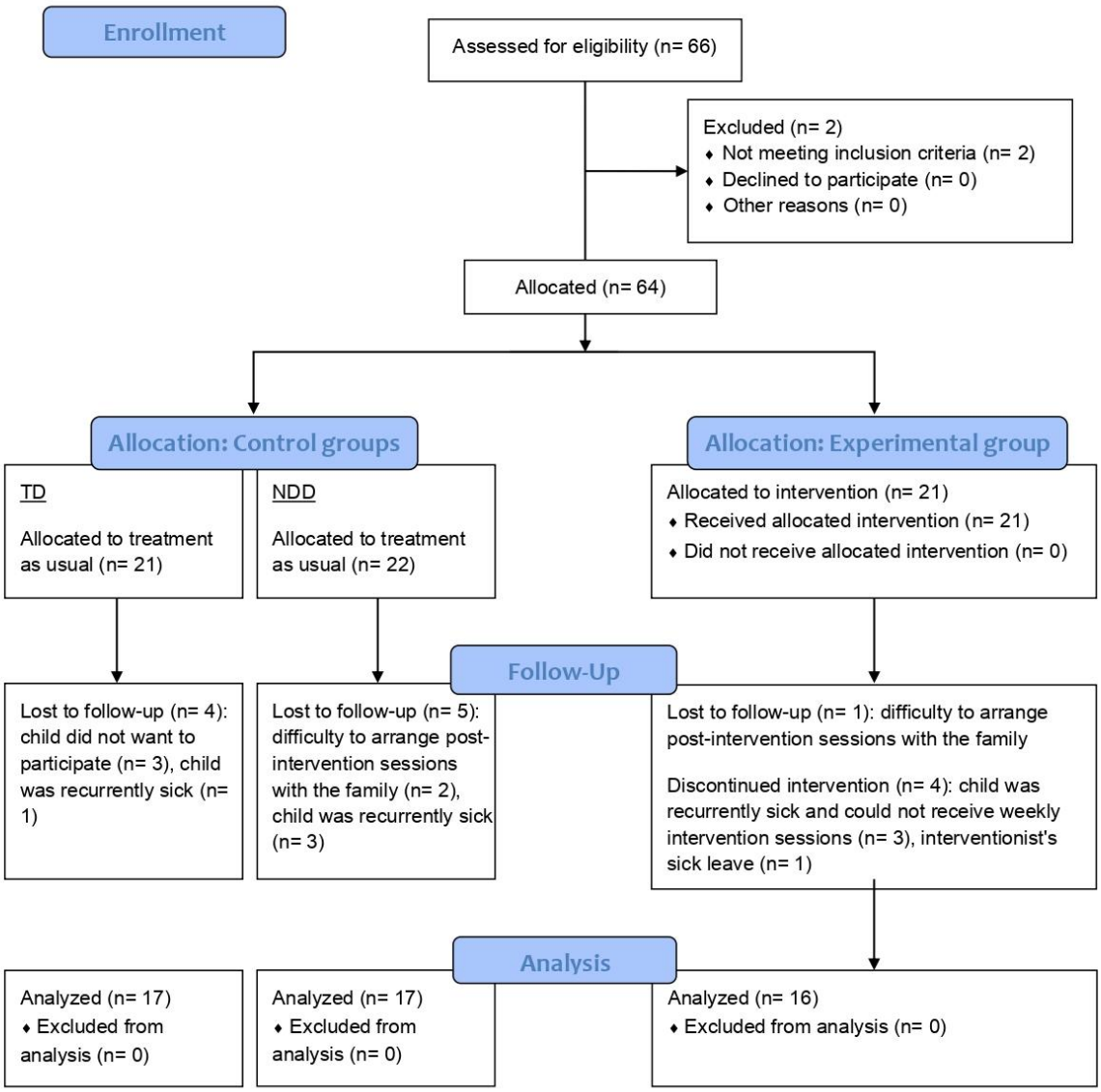

Adjusted R-squared = .502
Average of support needed to answer correctly
2
1
0
-1
2
3
4
5
6
Narrative macrostructure scores at post-test (untrained stories)

Appendices

A. Consort flow diagram

CONSORT
TRANSPARENT REPORTING of TRIALS

CONSORT 2010 Flow Diagram



B. Differences between participants with autism and Developmental Language Disorder: descriptive measures

	Control NDD		Experimental NDD	
Variable	**Autism**	**Developmental Language Disorder**	**Autism**	**Developmental Language Disorder**
N of participants (*N* of females and males)	4 (1 F, 3 M)	13 (4 F, 9 M)	8 (3 F, 5 M)	8 (3 F, 5 M)
N of participants with a diagnosis (*N* of F and M)	1 (1 F)	8 (1 F, 7 M)	6 (1 F, 5 M)	3 (3 M)
N of participants with reported risk (*N* of F and M)	3 (3 M)	5 (3 F, 2 M)	2 (2 F)	5 (3 F, 2 M)
Age: Mean (SD) Age: Range	4.65 (0.43) 4.17–5.08	4.78 (0.87) 3.92–7.08	4.63 (0.58) 3.50–5.25	5.78 (0.93) 4.75–7.33
TSI: Mean (SD) TSI: Range	87.70 (0) 87.70	97.32 (9.50) 87.7–110.8	99.64 (16.53) 70.5–120.2	105.23 (4.43) 101.1–110.6
CELF: Mean (SD) CELF: Range	95 (15.17) 73–106	79.23 (13.88) 58–100	85.75 (9.75) 66–99	77 (16.17) 58–105
PleaseApp: Mean (SD) PleaseApp: Range	20.00 (4.08) 14–23	19.62 (5.27) 10–30	26.00 (7.86) 19–38	22.00 (5.81) 13–29
K-BIT: Mean (SD) K-BIT: Range	101.75 (14.52) 86–116	101.62 (13.48) 74–123	105 (15.93) 82–127	98.13 (11.03) 83–110

C. Families' educational level

Out of the 50 participating children, 63 mothers, fathers and/or legal guardians reported information about their highest educational level. Please find the information below:

Educational level	**N (%)**
Obligatory Secondary Education	11 (17.46%)
A levels / High School diploma	15 (23.81%)
Vocational Education and Training	10 (15.87%)
Certificate of Higher Education	11 (17.46%)
Bachelor's degree	8 (12.70%)
Master's degree	5 (7.94%)
PhD	0 (0%)
Other	3 (4.76%)

D. Evaluation of professionals' treatment fidelity from videorecordings

1	From 1 to 7, did the professional follow the intervention procedure (1 being that she did not follow any activity and 7 that she followed all activities and in the appropriate order)? 1 2 3 4 5 6 7
2	From 1 to 7, did the professional use the recommended strategies of the intervention (e.g., verbal modeling, feedback, use of audiovisual support, and story enactment) (1 being that she did not use any strategy and 7 that she used all strategies appropriately)? 1 2 3 4 5 6 7

E. Narrative macrostructure and narrative perspective-taking coding rubrics

Narrative macrostructure (adapted from Demir et al., 2014)	
0	The retelling does not include any descriptive sequence.
1	The retelling includes one descriptive sequence (without any temporal sequence).
2	The retelling includes an action sequence (such as the main character and the problem).

3	The child produces an incomplete narrative that lacks two or more of the macrostructure elements (character, problem, attempt, solution, final).
4	The child produces an incomplete narrative that lacks one of the macrostructure elements (character, problem, attempt, solution, final).
5	The child produces a complete narrative that includes all macrostructure elements.
6	The child produces a complete narrative that includes all macrostructure elements and also adds details about the story.
Narrative perspective-taking (adapted from Dodd et al., 2011)	
0	The retelling does not include any emotion.
1	The retelling includes one emotion.
2	The retelling includes two or more emotions.
3	The retelling includes one emotion + its cause.
4	The retelling includes two or more emotions + the cause of at least 2 emotions.
+1	The retelling includes one mental term (such as *thinking*, *realizing, willing, wanting*).
+2	The retelling includes two or more mental terms (such as *thinking*, *realizing, willing, wanting*).

F. List of comprehension questions

Story	Questions
Story 1	1. Character. *Who was the main character?* 2. Initial emotion. *How did Meloix feel at the beginning?* 3. Problem. *Which problem did Meloix encounter?* 4. Emotion associated with problem. *How did Meloix feel after that?* 5. Cause of the emotion. *Why was Meloix angry?* 6. Attempt. *What does Meloix need to be clean?* 7. Resolution. *How did Meloix solve the problem?* 8. End. *How did the story end?* 9. Final emotion. *How did Meloix feel at the end of the story?*

	10. Cause of the emotion. *Why was Meloix happy at the end?*
Story 2	1. Character. *Who was the main character?* 2. Initial emotion. *How did Meloix feel at the beginning?* 3. Problem. *Which problem did Meloix encounter?* 4. Emotion associated with problem. *How did Meloix feel after that?* 5. First attempt. *How did Meloix try to solve the problem?* 6. Second attempt. *How did Meloix try to solve the problem the second time?* 7. Emotion associated with attempt. *How did Meloix feel when he could not* reach the bananas? 8. Third attempt. *How did Meloix try to solve the problem the thir time?* 9. Resolution. *How did Meloix solve the problem?* 10. End. *How did the story end?* 11. Final emotion. *How did Meloix feel at the end of the story?*
Story 3	1. Character. *Who were the main characters?* 2. Initial emotion. *How did Meloix feel at the beginning?* 3. First problem. *Which problem did Meloix encounter?* 4. Attempt. *How did Meloix try to solve the problem?* 5. Emotion associated to attempt. *How did Meloix feel after that?* 6. First resolution. *How did Meloix solve the problem?* 7. Second problem. *Which problem did Zepa encounter?* 8. Attempt. *How did Zepa try to solve the problem?* 9. Emotion associated to attempt. *How did Zepa feel after that?* 10. Second resolution. *How did Zepa solve the problem?* 11. Third problem. *Which problem did Meloix, Llampoc and Zepa encounter at the end?* 12. Emotion associated to problem. *How did they feel after that?* 13. Third resolution. *How did Meloix solve the problem?* 14. End. *How did the story end?* 15. Final emotion. *How did Meloix, Llampoc and Zepa feel at the end of the story?*

G. Descriptives (*M*, *SD* and *Range*) for pre- and post-intervention assessments

			Autism			
			Control NDD		**Experimental NDD**	
Variable			**PRE**	**POST**	**PRE**	**POST**
Narrative	**Trained**	*M*	2.00	1.50	2.25	3.63

macrostructure	**story**	*SD*	2.45	1.73	1.04	1.69
		Range	0–5	0–3	1–4	1–5
	Untrained stories	*M*	3.08	1.59	2.42	3.46
		SD	1.91	1.88	1.10	1.43
		Range	0.33–4.67	0–3.67	1–4	1.67–5.33
Narrative perspective-taking	**Trained story**	*M*	0.75	0	0	0.13
		SD	1.50	0	0	0.35
		Range	0–3	–	–	0–1
	Untrained stories	*M*	0.42	0	0.29	0.21
		SD	0.42	0	0.33	0.25
		Range	0–1	–	0–0.67	0–0.67

			Developmental Language Disorder			
			Control NDD		**Experimental NDD**	
Variable			**PRE**	**POST**	**PRE**	**POST**
Narrative macrostructure	**Trained story**	*M*	1.77	3.15	3.00	4.38
		SD	1.64	0.99	0.93	0.74
		Range	0–4	2–5	2–4	4–6

	Untrained stories	*M*	1.95	2.51	3.21	4.10
		SD	1.19	1.09	0.99	0.87
		Range	0–4	0.67–4.33	1.33–4	3–5.67
Narrative perspective-taking	**Trained story**	*M*	0.08	0.08	0	0.50
		SD	0.28	0.28	0	0.76
		Range	0–1	0–1	–	0–2
	Untrained stories	*M*	0.26	0.03	0.23	0.25
		SD	0.46	0.09	0.27	0.24
		Range	0–1.67	0–0.33	0–0.67	0–0.67