

1 **Assessing the effectiveness of multimodality in a classroom-based narrative intervention**

2

3

Júlia Florit-Pons, Alfonso Igualada, Pilar Prieto

4

5

**ABSTRACT**

6 This study assesses the [program] intervention aimed at enhancing children's oral narrative  
7 skills, focusing on its multimodal component, which includes story enactment by teachers and  
8 children, and an audiovisual multimodal narration by a storyteller. A cluster randomized trial  
9 was conducted (9 classrooms, 115 children), comparing a control group, a multimodal  
10 narrative-based group (complete [program]), and a narrative-based group ([program] without  
11 controlling for multimodality). The children's oral narrative macrostructure, microstructure,  
12 and perspective-taking skills were measured pre- and post-intervention. Results revealed that  
13 only the narrative-based group improved in narrative macrostructure and microstructure skills.  
14 No significant effects were found for the multimodal narrative-based group. Further analyses  
15 indicated that the children's active enactments involved in the multimodal component disrupted  
16 classroom dynamics, thereby preventing children's narrative learning. The study suggests that  
17 multimodality should be integrated in a way that complements, and does not disrupt, the  
18 narrative intervention activities.

19 **Keywords:** narrative intervention, multimodality, narrative macrostructure, narrative  
20 perspective-taking, narrative microstructure, preschool education

22 The development of the ability to narrate a story typically takes place during the preschool  
23 years and continues to progress throughout childhood, constituting one of the more complex  
24 abilities in child language development. Early narrative skills are considered as a reliable and  
25 ecologically valid measure for evaluating children's linguistic development (for a review, see  
26 Dickinson & McCabe, 2001). In addition, narrative skills have also been found to be both  
27 predictive and precursors of subsequent linguistic, literacy, social, and academic performance  
28 (e.g., Babayiğit et al., 2021; Dickinson & McCabe, 2001; Griffin et al., 2004). However,  
29 children naturally vary in their narrative abilities, highlighting the potential for narrative  
30 interventions to support language learning across different learners.

31 Narrative interventions allow language to be trained holistically, and researchers and  
32 practitioners have been devoted to developing interventions that promote oral narrative skills  
33 in preschool- and school-aged children. In the last years, different narrative-based interventions  
34 have been implemented and have generally been found to be effective (see Nelson & Khan,  
35 2019 and Pico et al., 2021 for reviews focusing on narrative interventions; for reviews on more  
36 general language-based interventions, refer to Donolato et al., 2023 and Walker et al., 2020).

37 Within the literature on narrative interventions, most studies have focused on boosting  
38 children's narrative macrostructure (i.e., the structure of the narrative discourse in terms of  
39 story grammar elements, such as character, problem, attempt, and resolution) and narrative  
40 microstructure (i.e., narrative productivity at the sentence level, such as the total number of  
41 words or the number of different words). Despite the variability in the duration of each narrative  
42 intervention (ranging from two-session interventions to whole academic year interventions),  
43 the vast majority of studies have reported positive outcomes after implementing the  
44 intervention for narrative macrostructure (e.g., Henry & Solari, 2020; Hettiarachchi, 2022),  
45 showing that training story grammar elements can also help improve narrative microstructure  
46 (e.g., Gillam, Olszewski et al., 2014; Spencer et al., 2018) (see e.g., Pico et al., 2021, for a  
47 review). These studies have also revealed that narrative interventions can promote  
48 generalization to other linguistic skills, such as vocabulary (Gillam, Olszewski et al., 2014;  
49 Suggate et al., 2021), or reading and writing (e.g., Spencer & Petersen, 2018b; Petersen et al.,  
50 2022). For instance, Spencer et al. (2018) implemented the registered narrative intervention  
51 *Story Champs* in three preschool classrooms throughout the school year and compared it to  
52 three other control classrooms. The authors conducted two narrative assessments in winter and

53 spring and found that, at both time points, the intervention group achieved significantly better  
54 story grammar skills (measured with a narrative retelling task) than the control group.  
55 Similarly, Pinto et al. (2019) compared 470 five-year-olds' narrative competence and reported  
56 that children who received the narrative intervention obtained significantly better narrative  
57 structure, coherence, and cohesion scores than those in control classrooms. For other studies  
58 reporting similar findings, see Daelman et al. (2024), Douglas et al. (2019), Gillam, Olszewski  
59 et al. (2014), Khan et al. (2014), Nelson et al. (2021), and West et al. (2021).

60 When telling or retelling a story, it is not only relevant to talk about what happened in the story,  
61 but also about how the characters felt and thought (Veneziano et al., 2020). Despite the  
62 significant effects on narrative measures, few studies have systematically trained and evaluated  
63 perspective-taking aspects within the story (i.e., the characters' emotions, perspectives, and  
64 mental states), but have generally considered emotions as another story grammar element (e.g.,  
65 Gillam, Gillam, & Laing, 2014; Spencer & Petersen, 2018a). One exception is the study by  
66 Veneziano et al. (2020), which implemented a short individualized narrative intervention with  
67 5- to 8-year-old French students. In this training session, the children were asked to retell a  
68 story by looking at a set of images, and then the interventionist started a conversation to talk  
69 about the character's mental states. The results showed that after the intervention, the narrative  
70 discourses of children receiving the intervention contained more mental states than those of  
71 children in the control group (see also Dodd et al., 2011 for students with autism). Within the  
72 context of preschool classrooms, Pronina and colleagues (2021) showed that a narrative-based  
73 intervention focusing on characters' emotions and mental states was effective in boosting  
74 expressive pragmatic skills (in terms of the appropriateness of answering different speech acts).  
75 These studies highlight the importance of incorporating direct instruction on emotions and  
76 perspectives into narrative interventions.

77 The effectiveness of most narrative interventions can be explained through the use of various  
78 educational strategies to support language learning. As reported in a recent review by Pico et  
79 al. (2021), most interventions provide specific instructions on unfolding narrative  
80 macrostructural elements, typically accompanied by visual support (e.g., story grammar icons,  
81 pictures, or storyboards). Another key strategy for training macrostructure elements is to  
82 employ verbal support, such as verbal prompting through question-and-answer sequences or  
83 positive feedback (e.g., repetitions and expansions). In line with these strategies, a recent study  
84 by Spencer and Petersen (2020) postulated a set of principles for effective narrative  
85 intervention, which included guidelines such as promoting active oral participation and making

86 the intervention as fun as possible, framing the stories in a socially relevant context for children,  
87 and providing different opportunities for children to tell and retell the whole story, which in  
88 turn allows for generalization to other stories. To the best of our knowledge, there is an  
89 educational strategy that has been incorporated less frequently into narrative interventions,  
90 namely multimodality.

91 Multimodality refers to how we use our body, hands, face, and voice to communicate (Perniss,  
92 2018). A growing body of research has shown that multimodal communicative behaviors, such  
93 as gestures and facial expressions, play a foundational role in language development. Early  
94 multimodal skills are predictive of later oral language, vocabulary, and narrative competence,  
95 suggesting that children's multimodal abilities scaffold their acquisition and organization of  
96 language (for reviews, see Esteve-Gibert & Prieto, 2018; Hostetter, 2011; Kirk et al., 2022;  
97 Vilà-Giménez & Prieto, 2021). Understanding these links provides a rationale for incorporating  
98 multimodal strategies into narrative interventions, as they may offer an additional pathway for  
99 supporting language learning.

100 In narrative storytelling, multimodality refers to enactment strategies such as using gestures to  
101 represent concepts like events or actions (e.g., Demir et al., 2015; Li, 2024; Macoun & Sweller,  
102 2016), having a storyteller narrate and enact the story (e.g., Macoun & Sweller, 2016; Vilà-  
103 Giménez et al., 2019), or promoting story enactment, either by the interventionist or the  
104 children themselves (e.g., Bernstein et al., 2024; Vilà-Giménez & Prieto, 2020). In fact, lab-  
105 based studies have found that observing gestures during a storyteller's narration enhanced  
106 children's narrative recall and retelling skills (e.g., Guilbert et al., 2021; Macoun & Sweller,  
107 2016; McKern et al., 2021; Vilà-Giménez et al., 2019). In fact, it is not only observing gestures  
108 that helped, but also producing them while narrating the story again (e.g., Parrill et al., 2018;  
109 Vilà-Giménez & Prieto, 2020). Vilà-Giménez and Prieto (2020) showed that a one-session  
110 individualized narrative intervention involving both observing and producing gestures was  
111 effective for promoting five-year-old children's narrative macrostructure skills. All of these  
112 studies are in line with the multimodal enrichment paradigm (Mathias & von Kriegstein, 2023),  
113 which claims that our body can promote language learning (see also Dargue et al., 2019;  
114 Goldin-Meadow, 2014; Hostetter, 2011; Rohlfing, 2017; Vilà-Giménez & Prieto, 2021 for  
115 reviews on the effects of gesture use on comprehension and production). However, it should  
116 be noted that these abovementioned lab-based studies tend to focus mainly on hand gestures,  
117 with less emphasis on broader forms of multimodality involving other articulators.

118 Based on this, a handful of recent studies have gone one step further and have tested the effects  
119 of multimodality in real-life classrooms, incorporating not only hand gestures but also broader  
120 forms of enactment such as whole-body enactment and dramatization (e.g., Bernstein et al.,  
121 2024; Murachver et al., 1996; Nicolopoulou et al., 2015). First of all, to the best of our  
122 knowledge, only a few of classroom-based narrative interventions have systematically  
123 incorporated a multimodal component, with mixed findings. One example is the *Story Champs*  
124 intervention (Spencer & Petersen, 2018a) for preschool- and school-aged children, where  
125 conventional or metaphoric gestures are used to represent abstract concepts, such as the  
126 different macrostructural elements (e.g., placing the fingertips together to represent the setting  
127 of the story). Although the specific contribution of these gestures within *Story Champs* has not  
128 been tested, significant narrative improvements after the intervention have been found (e.g.,  
129 Petersen et al., 2022; Spencer et al., 2018). Other studies have used a more naturalistic approach  
130 involving child-directed storytelling and story enactment activities, and have directly tested its  
131 effects. For instance, Pronina et al. (2021) systematically incorporated multimodality into a  
132 narrative intervention using facial expressions to act out different emotional and mental terms.  
133 The authors compared the performance on pragmatic and socioemotional skills of three  
134 different groups of children: a control group not receiving the intervention, a group receiving  
135 the intervention with the systematic multimodal embodiment of facial expressions, and another  
136 group receiving the intervention without the multimodal component. The results showed that  
137 both groups receiving the intervention (with no difference between them) improved their  
138 expressive pragmatic skills, but that the intervention was not effective in improving emotion  
139 understanding and mental state comprehension. More broadly, other studies have incorporated  
140 story enactment strategies as drama-based activities using pantomime gestures, which convey  
141 a narrative or story in the absence of speech (see e.g., Bernstein et al., 2024; see also Corbett  
142 et al., 2016 for autistic children and Nicolopoulou et al., 2015 for children from households  
143 with a low income; for reviews on drama-based interventions, see Frydman & Mayor, 2024  
144 and Lee et al., 2015). Recent review studies have generally found that drama-based  
145 interventions can support a range of developmental outcomes in children, particularly  
146 enhancing narrative skills, oral language, and socioemotional understanding, although effects  
147 vary depending on the type, intensity, and implementation fidelity of the intervention (Frydman  
148 & Mayor, 2024; Lee et al., 2015). For instance, Ionescu and Illie (2018) showed that the group  
149 preschool children that had been exposed to an embodied narrative intervention (by  
150 dramatizing the story events with story-related objects) were better in retelling the story and in  
151 remembering key vocabulary and idioms than the group of children that had been exposed to

152 only listening to the story. Also, the recent study by Bernstein and colleagues (2024) compared  
153 children's narrative recall abilities after receiving a classroom-based storybook reading  
154 intervention either in the absence or presence of a drama-based activity to act as the main  
155 character. Their results showed that, although there were no differences between groups in the  
156 quality of children's story retell, those receiving the intervention used more gestures when  
157 retelling the stories.

158 Overall, with all these studies, we have seen the value of oral narrative interventions for  
159 children's language development. However, the specific contribution of multimodality to  
160 narrative intervention practices is less clear. Although studies conducted in an individualized  
161 or lab-based context have reported positive effects of multimodality, the only two studies  
162 testing the effects of multimodality with groups of typically developing students in classrooms  
163 (e.g., Bernstein et al., 2024, Pronina et al., 2021) have shown null results. To address the  
164 conflicting evidence between the positive effects of multimodality found at the more individual  
165 one-to-one lab-based and the lack of effects at the classroom level, we believe that further  
166 research is needed to rigorously evaluate the effectiveness of this multimodal component in  
167 real-world classroom contexts, where findings could inform broader educational practice.  
168 Moreover, provided that teachers inherently use gestures in their professional practice, we need  
169 more research on how multimodality should be systematically incorporated into narrative  
170 intervention practice, and its potential effectiveness should be further investigated. This  
171 approach would enable teachers to not only implement this strategy effectively in the classroom  
172 but also acknowledge its potential value (e.g., Seccia & Goldin-Meadow, 2024).

173 The present study aims to assess the effectiveness of the novel multimodal narrative-based  
174 intervention (the [program]; blinded, 2025a), and particularly, the potential effectiveness of its  
175 core multimodal component. The [program] was designed for implementation in preschool  
176 classrooms as a supplementary component of standard curricular activities. The intervention  
177 systematically trains narrative macrostructure and perspective-taking elements with the aid of  
178 validated educational strategies, together with a strong multimodal component that affects the  
179 behavior of the three social actors involved in the intervention, namely the enactment by the  
180 teacher, the enactment by children, and the observation of an audiovisual multimodal narration  
181 by a storyteller. The present study specifically seeks to test whether the [program] intervention  
182 is effective in improving preschool children's oral narrative skills (i.e., macrostructure,  
183 microstructure, and perspective-taking). To do this, children's narrative performance after a  
184 full implementation of [program] for three weeks will be compared to a group of children

185 receiving the same [program] intervention without the multimodal component and to a group  
186 of children not receiving the intervention. Children's narrative performance will be evaluated  
187 using a narrative telling task with two untrained stories, to see whether narrative gains are  
188 observed in novel but structurally similar materials. We hypothesize that children who receive  
189 the intervention with a systematic multimodal component will show greater improvements than  
190 those who receive the same narrative-based intervention without the controlled multimodal  
191 component (e.g., Mathias & von Kriegstein, 2023) and those that do not receive the intervention  
192 (e.g., Nelson & Khan, 2019; Pico et al., 2021). This hypothesis builds on the existing evidence  
193 on the causal role of multimodality with the idea that narrative learning is supported not only  
194 by the oral language activities involved in the intervention, but also by engaging with it through  
195 the body, provided that multimodal strategies can allow children to experience the events of  
196 the story, helping them to internalize and 'live' the narrative from a first-person perspective.

EN REVISIÓ

**198 Design**

199 To test the effectiveness of the [program], particularly the contribution of its multimodal  
200 component, this study followed a cluster randomized controlled trial with three different  
201 groups: a control group that did not receive the intervention and two experimental groups. One  
202 experimental group received the [program] intervention (henceforth multimodal narrative-  
203 based group) and another one received the [program] intervention without the multimodal  
204 component (henceforth narrative-based group). The study involved three participating schools,  
205 each contributing three classrooms. Group assignment was done at the classroom level (i.e.,  
206 entire classrooms within each school, rather than individual students, were assigned to groups),  
207 with each group consisting of three classrooms that were randomly allocated to one of the three  
208 groups.

**209 Participants**

210 A total of 115 children (55 girls) from an initially recruited sample of 136 children participated  
211 in this study (see Appendix I for the CONSORT 2010 flow diagram of the process from  
212 participant enrollment until analysis). The sample size for this study was calculated *a priori*  
213 using G\*Power Version 3.1.9.6 (Faul et al., 2009) with inputs of .05 for alpha, three groups,  
214 and two for the number of measurements. The sample size necessary to have an acceptable  
215 level of power ( $1-\beta$  error probability = 0.8) and to detect small effect sizes of  $f=0.15$  (Cohen,  
216 1988) was 111 participants.

217 Participants were Spanish-Catalan bilinguals<sup>1</sup> aged between 4;9 and 6;2 attending their last  
218 year of preschool. Participants were recruited from nine preschool classes in three public  
219 schools in L'Hospitalet de Llobregat, a city in the metropolitan area of Barcelona, Catalonia.  
220 All three schools were located in the same neighborhood, with middle-low to middle  
221 socioeconomic status (according to the socioeconomic territorial index calculated by the  
222 Statistical Institute of Catalonia). The inclusion criteria were as follows: children attending the  
223 last year of preschool (I5 in the Catalan education system), with the only restriction that they  
224 needed to have normal hearing. The final sample consisted of 106 children with typical

---

<sup>1</sup> Although all children were bilingual, Catalan is the vehicular language and the language of instruction in their schools. Therefore, all assessment and intervention activities were carried out in Catalan, and bilingual status was not a variable of interest in the present study or a factor in the study design.

225 development, while 9 children had a professional report indicating developmental difficulties  
 226 with language and communication. Participants' structural language abilities and non-verbal  
 227 IQ were evaluated using standardized tests before the intervention (K-BIT by Kaufman &  
 228 Kaufman, 1990, for non-verbal IQ; and CELF-5 by Wiig et al., 2013, for structural language).  
 229 A set of pairwise t-tests was performed to assess potential differences between the three groups.  
 230 No significant differences were found between the groups in terms of age ( $ps > .361$ ), structural  
 231 language ( $ps > .135$ ), or non-verbal IQ ( $ps > .467$ ). See Table 1 below for the descriptives of  
 232 the participants' demographic, linguistic and cognitive characteristics.

233 **Table 1.** Participants' demographic, linguistic and cognitive characteristics

	<b>Control</b>	<b>Narrative-based</b>	<b>Multimodal narrative-based</b>
<i>N</i> of participants ( <i>N</i> of females and males)	38 (16 F, 22 M)	41 (20 F, 21 M)	36 (19 F, 17 M)
<i>N</i> of participants with language difficulties ( <i>N</i> of F and M)	4 (2 F, 2 M)	2 (2 M)	3 (3 M)
Age: Mean (SD)	5.38 (0.33)	5.40 (0.35)	5.45 (0.38)
Age: Range	4.75–6.08	4.75–6.17	4.75–6.17
CELF-5: Mean (SD)	89.14 (16.99)	92.76 (16.19)	87.29 (13.91)
CELF-5: Range	40–115	55–120	55–113
K-BIT: Mean (SD)	101.58 (12.73)	102.73 (12.46)	100.57 (13.62)
K-BIT: Range	72–123	77–141	66–123

234  
 235 This study received ethical approval from the Institutional Committee for Ethical Review of  
 236 Projects at the host university of the first and third authors (2021, ref.: 228) and from the  
 237 Regional Ministry of Education (Catalan Government). Legal guardians were provided with  
 238 detailed information about the project and provided written consent for their child to participate  
 239 in the study. Given that the intervention was implemented as a regular class activity, the legal  
 240 guardians of the children allocated to the intervention groups were informed that if they did not  
 241 provide consent for their child to participate, their child would still participate in the

242 intervention sessions, but they would not be administered the pre- and post-intervention  
243 assessments.

## 244 **Pre- and post-intervention assessment**

### 245 *Materials*

246 To determine whether the intervention was effective in boosting children's oral narrative skills  
247 (i.e., macrostructure, microstructure and perspective-taking), children completed a narrative  
248 telling task at both pre and post-intervention assessments, consisting of two untrained stories.  
249 The inclusion of untrained stories allowed us to assess children's generalization to novel but  
250 structurally similar materials.

251 The two stories were animated wordless cartoons from the series *Die Sendung mit der Maus*,  
252 each lasting approximately 50 seconds. At pre-intervention, children first narrated a story  
253 featuring the main character, a mouse, and then a second story featuring two characters: the  
254 mouse and a small elephant. At post-intervention, children narrated two new stories featuring  
255 the same characters (i.e., the mouse and the elephant). This ensured the stories were novel at  
256 each assessment and could therefore serve as a test of generalization rather than memory or  
257 repeated exposure. At both time points, the stories were administered in a fixed order: first the  
258 one-character story, and then the two-character story.

259 Despite the difference in the number of characters, both stories shared similar macrostructural  
260 elements in the logical order, which allowed for a meaningful comparison of narrative skills  
261 across stories. To ensure this similarity, the authors analyzed the stories by comparing key  
262 narrative components, including the number and order of problems, attempts, and resolutions,  
263 as well as the logical sequence of these macrostructural elements. In addition, emotional and  
264 mental states were examined at critical points in each story –at the beginning, following the  
265 problem, and as a consequence of the resolution. Across all stories, this analysis confirmed a  
266 comparable episodic organization characterized by a single problem, a goal-directed attempt to  
267 resolve it, a clear outcome, and consistent placement of emotional and mental state cues. This  
268 ensured that the stories were similar in their overall organization, despite differences in the  
269 number of characters.

### 270 *Procedure*

271 The task was administered to each child individually in a silent room at the school during school  
272 hours. First, the child watched the cartoon, and the experimenter then asked him/her to stand  
273 up and narrate the story. The child was asked to stand up with the aim that their gestures and  
274 enactment of the story could be further investigated in a follow-up study, and to encourage a  
275 more natural use of multimodal expression and gesture. The task was presented in the form of  
276 a game, such that after the child narrated the story, the experimenter had to guess –from a set  
277 of four still images– the story the child was talking about. This process was repeated for the  
278 second story, always following the same order: first the story featuring one character, and then  
279 the story featuring two characters. The task was administered before and after the intervention  
280 (i.e., between one or two weeks after the last intervention session). All pre- and post-  
281 intervention sessions were video recorded for them to be revised and coded after the sessions.

## 282 *Coding*

283 Each narration was coded for narrative macrostructure, microstructure and perspective-taking.  
284 While macrostructure and perspective-taking were coded during the data collection by the  
285 experimenter (i.e., a research assistant), narrative microstructure was coded *a posteriori* by the  
286 first author together with two other predoctoral researchers and a research assistant. Real-time  
287 coding was selected for practical and methodological reasons. Given the brief and simple nature  
288 of the stories, coders were able to identify macrostructural and perspective-taking elements  
289 reliably as children narrated. All coders were graduate-level research assistants who had  
290 received specific training from the first author using sample narratives until full familiarity  
291 with the coding schemes was reached, allowing them to easily code for macrostructure and  
292 perspective-taking at the same time. During data collection, all sessions were video recorded,  
293 allowing coders to review any uncertain cases and revise their ratings when necessary.

294 First, narrative macrostructure was coded using a 0–6 coding scheme adapted from Demir et  
295 al. (2014) to assess children’s introduction of different macrostructural elements, with the score  
296 increasing depending on the number of story elements introduced. For instance, a score of 6  
297 indicated that the child introduced all elements and added extra details about what happened  
298 during the plot, whereas a score of 0 indicated that the child did not introduce any elements or  
299 a descriptive sequence. To see the full coding scheme, refer to Appendix II.

300 Narrative microstructure involved two narrative productivity measures, namely the total  
301 number of words (TNW, also known as *tokens*) and the total number of different words  
302 (TNDW, also known as *types*). After all narrations were transcribed following the CHAT

303 guidelines (MacWhinney, 2000), TNW and TNDW were automatically calculated using the  
304 CLAN software (MacWhinney, 2000). The coding for narrative microstructure could only be  
305 performed on 107 participants (rather than all 115), provided that eight participants (three from  
306 the control group, two from the narrative-based group and three from the multimodal narrative  
307 group) had to be excluded because the narration could not be video recorded due to technical  
308 difficulties or because there was too much background noise, and therefore the narration could  
309 not be annotated.

310 Narrative perspective-taking was coded using a 0–6 coding scheme adapted from Dodd et al.  
311 (2011), who assessed children’s ability to introduce emotional and mental terms within the  
312 narration (e.g., *angry*, *happy*, *sad*, *surprised*, and *thinking*, *willing*, *realizing*), with the score  
313 increasing depending on the number of elements introduced as well as the relation among them.  
314 A score of 6 corresponded to a narration that introduced at least two emotional terms, the cause  
315 of these emotions, and at least two mental terms, while a score of 0 corresponded to a narration  
316 that did not include any emotional or mental term. The full coding scheme is described in  
317 Appendix III.

318 Overall, the narrative measures collected in this study are well-established and valid indicators  
319 of children’s narrative competence. The coding schemes for narrative macrostructure and  
320 perspective-taking were adapted from existing studies (Demir et al., 2014; Dodd et al., 2011),  
321 and the microstructure measures (TNW and TNDW) are standard indices of narrative  
322 productivity widely used in the literature (MacWhinney, 2000). Together, these measures  
323 provide a comprehensive and ecologically valid assessment of children’s ability to organize  
324 story content, use diverse vocabulary, and integrate emotional and mental-state information in  
325 their narratives.

### 326 *Inter-rater reliability*

327 Provided that narrative macrostructure and narrative perspective-taking codings were  
328 perceptual assessments, inter-rater reliability was calculated using data from 20 participants,  
329 corresponding to 80 stories (2 stories × 2 time points) and 17.39% of the data. The participants  
330 were randomly selected from the whole set of participants using the website [www.random.org](http://www.random.org),  
331 with no particular distribution for schools or classes. Reliability was calculated using the  
332 original scores and the scores of a trained research assistant. Cohen’s kappa (weighted kappa)  
333 was used to calculate inter-rater reliability. The results for narrative macrostructure showed  
334 83.8% agreement among coders and a Cohen’s kappa of .899. Regarding narrative perspective-

335 taking, the results indicated 97.5% agreement and a Cohen's kappa of .861. Overall, the kappa  
336 value for both measures showed moderate agreement, as per McHugh (2012).

### 337 **Intervention phase**

#### 338 *The [program]*

339 The [program] is a multi-tiered narrative-based intervention developed in a co-creation fashion  
340 with more than 90 preschool teachers and speech-language therapists working in Catalan  
341 schools and speech therapy centers (blinded, 2025b). In this study, we used the universal tier  
342 (tier 1) of the [program], which was implemented at the classroom level.

343 *Intervention protocol.* The [program] aims to train oral narrative macrostructure and  
344 perspective-taking skills through interaction activities and crucially involves a dedicated  
345 multimodal component. The intervention protocol revolves around three different wordless  
346 cartoons from the Colombian series *Chigüiro*, each of which is trained for three consecutive  
347 days (three stories × three sessions = nine intervention sessions). The protocol is as follows: all  
348 intervention sessions follow the same initial set of activities: introduction to the session,  
349 watching the wordless cartoons, a short teacher-children interaction about the story plot, and  
350 watching a video of the storyteller narrating the story. After these activities, each session has a  
351 central activity. In the **first session** of each story, the central activity consists of a sequential  
352 enacted retelling using a question-and-answer sequence with the aid of story icons and short  
353 animated videos depicting the main elements and events of the story. The procedure is as  
354 follows: the teacher asks a question about each story element (e.g., *Which problem did the*  
355 *character encounter?*), the children answer (e.g., *Meloix was dirty.*) and she then gives  
356 feedback by providing a full sentence model (e.g., *Meloix was dirty because he was playing*  
357 *with his ball and the ball fell on the mud.*). While providing this model, the teacher enacts the  
358 action and after that, encourages all children to stand up and also enact it. This process is  
359 repeated for all elements of the story (i.e., character, problem, attempt, resolution, end,  
360 emotions). In the **second session** of each story, the central activity centers around the retelling  
361 and dramatization of the story. While children remain seated, the teacher asks a question  
362 targeting all story elements. For each question, two children are invited to the front of the class:  
363 one takes on the role of the storyteller and responds to the question, while the other takes on  
364 the role of the actor and acts out the corresponding action or emotion. The teacher then provides  
365 positive feedback after the performance. This process is repeated for all story elements with  
366 different pairs of children, with all pairs participating at least once. Once all questions have

367 been answered, children are invited to volunteer to retell and dramatize the entire story in pairs  
368 in front of the class. The central activity during the **third session** of each story focuses on the  
369 generation of personal stories. First, the teacher tells and enacts a personal narrative related to  
370 the trained fictional story and encourages children to stand up in front of their classmates and  
371 tell their personal stories. The intervention materials included written instructions specifying  
372 when teachers should enact each story element and encourage children to also do it. However,  
373 they were not given any strict instructions on how to perform the enactment; rather, they were  
374 asked to enact the elements in their own natural manner.

375 The intervention was implemented over three consecutive weeks, involving three  
376 weekly sessions (implemented each Monday, Wednesday and Friday), each of which lasted  
377 approximately 25 to 30 minutes. All intervention sessions were implemented in Catalan, the  
378 vehicular language in schools in Catalonia, and by each classroom teacher ( $n = 6$ ), for it to be  
379 as natural as possible for children receiving the sessions.

380 For more detailed information on the [program] intervention protocol, please refer to  
381 blinded (2025a).

382 *Differences between the two experimental groups.* As mentioned above, two different  
383 groups received the same [program] intervention, with the exception that one group received it  
384 with the presence of the multimodal component (multimodal narrative-based), the other  
385 received it without the multimodal component (narrative based).

386 The multimodal component is integrated into the intervention involving the three social  
387 actors of the intervention, namely, the audiovisual recordings of the storyteller, the teacher, and  
388 the children. First, we have the video of the storyteller explaining the story: children in the  
389 multimodal narrative-based group watched the video of the storyteller in which she used not  
390 only child-directed prosody but also child-directed enactment strategies, implemented through  
391 the use of different gestures and facial expressions. The second social actor is the teacher, who  
392 apart from using natural multimodal strategies during the intervention, received scripted  
393 prompts to enact the main macrostructural and perspective-taking elements of the stories.  
394 Finally, the third actor refers to the children receiving the intervention: teachers also had  
395 instructions to encourage the children to stand up and also actively enact the macrostructural  
396 and perspective-taking elements of the stories. As for the narrative-based group, the video of  
397 the storyteller included a narration in which the storyteller was naturally telling the story  
398 without using gestures, but maintaining natural child-directed facial expressions and a child-

399 directed prosodic style. In addition, teachers in the narrative-base group did not have any  
400 specific instruction to enact the stories, but were rather asked to act naturally. Children were  
401 not encouraged to enact the story and thus remained seated throughout the session. Overall, all  
402 macrostructural elements (except for the character) were enacted using iconic gestures that  
403 embodied the main actions, such as bouncing a hand toward the floor to show a ball falling into  
404 mud (problem), or jumping with a raised hand to represent reaching for bananas in a tree  
405 (attempt). Perspective-taking elements were enacted using mostly facial expressions to enact  
406 emotions but also some metaphorical gestures (such as the thinking hand gesture) to enact  
407 mental terms. Figure 1 shows a visual representation of the differences between the two  
408 experimental groups regarding the multimodal component, involving the video of the  
409 storyteller, the teacher and the children themselves.

410 [FIGURE 1 HERE]

411 **Figure 1.** Schematic representation of the implementation of the multimodal component. Left  
412 panel shows the multimodal narrative-based group and the right panel shows the narrative-  
413 based group.

414 To ensure that the videos of the storyteller were comparable across the two groups, a group of  
415 eight preschool teachers (unfamiliar with the intervention and blind to the intervention group)  
416 evaluated the videos on a Likert scale ranging from 1 to 7. It was found that both stimuli were  
417 similar in terms of how natural the narration was (narrative-based:  $M = 5.78$ ,  $SD = 1.12$ ;  
418 multimodal narrative-based:  $M = 6.44$ ,  $SD = 0.85$ ).

419 *Intervention fidelity.* The intervention fidelity was assessed using two complementary  
420 measures, e.g. teachers' self-assessment and external *a posteriori* evaluation by the researchers.  
421 First, after each intervention session, all teachers were asked to complete a short questionnaire  
422 that reported the duration of the session and the degree to which they had followed the  
423 intervention protocol. The average duration of the sessions was 27.93 minutes ( $SD = 4.87$ ),  
424 with a similar duration for both groups (narrative-based:  $M = 28.12$ ,  $SD = 6.10$ ; multimodal  
425 narrative-based:  $M = 27.78$ ,  $SD = 3.86$ ). Teachers reported that 95% of the time they followed  
426 the intervention protocol (narrative-based: 91.30%; multimodal narrative-based: 100%).

427 Additionally, intervention fidelity was evaluated *a posteriori* through video recordings  
428 of the intervention sessions. A research assistant blinded to the intervention group watched a  
429 total of 18 intervention sessions (i.e., three sessions  $\times$  six teachers), corresponding to 33% of

430 the data, and assessed the teacher's treatment fidelity using a 1-to-7 Likert scale (1 = the teacher  
431 did not follow the intervention protocol and 7 = the teacher followed the protocol step by step).  
432 The results showed that the interventionists systematically followed the intervention procedure  
433 ( $M = 6.33$ ,  $SD = 0.84$ ; narrative-based:  $M = 5.67$ ,  $SD = 0.71$ ; multimodal narrative-based:  $M =$   
434  $7$ ,  $SD = 0$ ). The coder was also asked to assess the implementation of multimodality by the  
435 teachers (i.e., how expressive they were with their body using different bodily gestures and  
436 facial expressions). Teachers in both groups used gestures and facial expressions throughout  
437 the sessions (on a scale of 1 to 7, multimodal narrative-based:  $M = 6.56$ ,  $SD = 0.73$ ; narrative-  
438 based:  $M = 3.33$ ,  $SD = 1$ ). This was because all teachers (both in the narrative-based group and  
439 the multimodal narrative-based group) were asked to act naturally, with the only difference  
440 being that the teachers in the multimodal group were given specific instructions on how to use  
441 gestures throughout the session.

442 A complementary assessment of the implementation of the intervention was also  
443 conducted *a posteriori* with the aim to further investigate potential classroom disruptive  
444 behaviors across the groups.

445 *Control group: Treatment as usual*

446 Finally, the three class groups constituting the control group did not receive any intervention  
447 but continued with their usual sessions involving standard curricular activities. Detailed *a*  
448 *posteriori* reports from the three preschool teachers leading the control groups explained that  
449 during the target weeks in which the other groups were taking the interventions they were  
450 involved in educational activities to promote new vocabulary and phonological awareness, to  
451 establish routines, and to initiate mathematical thinking. No oral narrative activities were  
452 performed in any control group.

### 453 **Statistical analyses**

454 To assess the effectiveness of the intervention on children's oral narrative skills, four Linear  
455 Mixed-Effects (LME) models were performed using the *lme4* package (Bates et al., 2015) in  
456 R. The score for each narrative measure (i.e., macrostructure, microstructure (TNW and  
457 TNDW) and perspective-taking), which was the average score of the two narrations, was  
458 established as the dependent variable. All models included Test (two levels: pre- and post-  
459 intervention) and Group (three levels: control, narrative-based, multimodal narrative-based) as  
460 fixed factors, together with their two-way interaction. The best random-effects structure was

461 determined using the *performance* package (Lüdtke et al., 2024). Candidate models included  
462 varying intercepts for participants, language score, classrooms, and schools, either alone or in  
463 combination. The final models included varying intercepts by language and by participant, or  
464 varying intercepts by language only. Finally, for all significant main effects and interactions,  
465 post-hoc pairwise comparisons were performed using the Bonferroni correction with the  
466 *emmeans* package (Lenth, 2021) and a measure of effect size (via Cohen's *d*). For significant  
467 interactions, we examined the data in two complementary ways: (1) by assessing within-group  
468 changes across time (pre- to post-intervention) and (2) by comparing group differences at post-  
469 intervention. This approach offers a more fine-grained understanding of the intervention effects  
470 and helps to disentangle whether observed changes stem from improvements within groups,  
471 differences between groups, or both.

472 To assess potential differences between the target variables, we conducted equivalence tests  
473 using the *TOSTER* package (Lakens, 2017) to better interpret any null findings that may come  
474 up from the inferential statistical analyses. For the macrostructure and perspective-taking  
475 measures, we defined the smallest effect size of interest as  $\pm 0.5$  raw points, as it reflects  
476 minimal, non-meaningful variability based on the measurement characteristics. For TNW and  
477 TNDW, which are count variables without a fixed range, we used ratio-based equivalence  
478 bounds rather than raw-unit differences, as these are appropriate for linguistic count data  
479 (Winter, 2020). We defined the smallest meaningful difference as a 20% change (i.e., a 0.80–  
480 1.25 ratio). A significant TOST result indicates that the 90% confidence interval falls entirely  
481 within these bounds and thus supports the interpretation of equivalence; whereas a non-  
482 significant TOST result indicates that the equivalence cannot be concluded.

483 Particularly, the analyses for narrative macrostructure and narrative perspective-taking were  
484 performed using data from all 115 participants, whereas the analyses for narrative  
485 microstructure were performed with data from 107 participants (see *Coding* subsection for  
486 more details).

487 As a follow-up to the general assessment, we descriptively examined the narrative skills of  
488 children identified as having language and communication needs to explore whether they  
489 showed improvements after receiving only tier 1 instruction. Given the small sample size ( $N =$   
490 9), no inferential statistics were conducted. Instead, we calculated descriptive statistics to  
491 observe patterns of change across conditions as well as gains compared to the typically  
492 developing peers.

493

## RESULTS

494 This section describes the results obtained from the different analyses. We first provide  
495 descriptive statistics for all narrative measures, and then report the statistical results of the LME  
496 models conducted with each narrative measure.

### 497 **Descriptives**

498 Table 2 shows the descriptive statistics for each narrative measure at pre- and post-intervention  
499 assessments broken down by group (control, narrative-based, multimodal narrative-based).

EN REVISIÓ

**Table 2.** Descriptives for pre- and post-intervention narrative assessments

	Value	Control		Narrative-based		Multimodal narrative-based	
		PRE	POST	PRE	POST	PRE	POST
<b>MACRO.</b>	M (SD)	3.86 (1.25)	3.64 (1.13)	3.65 (1.16)	4.02 (0.90)	3.96 (0.77)	4.08 (0.99)
	Observed range	0–5.50	0–5.50	0–5.50	2–5.50	2–5.50	1–5.50
<b>MICRO.: TNW</b>	M (SD)	40.96 (16.03)	36.86 (15.80)	37.88 (13.19)	46.65 (14.33)	42.40 (12.72)	42.74 (14.28)
	Observed range	5–71	18–78	12–64	16–74	18–68	10–65
<b>MICRO.: TNDW</b>	M (SD)	26.03 (7.73)	23.97 (7.13)	24.36 (6.70)	28.28 (6.74)	25.74 (5.88)	25.89 (6.29)
	Observed range	4–40	15–42	10–39	13–42	15–39	9–39
<b>PERS.</b>	M (SD)	0.17 (0.37)	0.05 (0.16)	1.16 (0.25)	0.90 (0.12)	0.29 (0.35)	0.17 (0.40)
	Observed range	0–1.50	0–0.50	0–1	0–1.50	0–1	0–2

501 *Note 1.* MACRO. = macrostructure; MICRO. = microstructure; TNW = total number of words; TNDW = total number of different words; PERS.  
502 = perspective-taking; PRE = pre-intervention assessment; POST = post-intervention assessment

503 *Note 2.* The possible range for narrative macrostructure and narrative perspective-taking was 0–6. The possible range for the narrative  
504 microstructure measures was 0–NA, with NA indicating that no theoretical maximum value is defined for this measure.

505 *Note 3.* The score corresponds to the average score of the two narrations.

506 **Narrative macrostructure**

507 The LME model assessing narrative macrostructure only showed significant results for the two-  
508 way interaction between Test and Group. The post-hoc pairwise comparisons reported that only  
509 the narrative-based group significantly improved from pre- to post-intervention ( $d = .56, p =$   
510  $.013$ ). In addition, no significant differences were observed between groups at post-  
511 intervention. For the full statistical results report, see Table 3 (see also Figure 2).

512 To further interpret the non-significant differences observed in the control and multimodal  
513 narrative-based groups for the macrostructure scores, we conducted equivalence tests  
514 comparing pre- and post-intervention scores within each group using  $\pm 0.5$  points as the  
515 equivalence bounds. For the control group, the equivalence test was not significant ( $p = .876$ ),  
516 and the 90% CI for the mean change ( $-0.18, 4.30$ ) exceeded the  $\pm 0.5$  equivalence bounds;  
517 therefore, pre–post scores could not be considered equivalent. For the multimodal group, the  
518 equivalence test was significant ( $p = .011$ ), with the 90% CI for the mean change ( $-0.39, 0.14$ )  
519 falling within the  $\pm 0.5$  bounds. Thus, pre–post scores in the multimodal group were statistically  
520 equivalent. To further examine the non-significant group differences at post-intervention, we  
521 conducted equivalence tests between all three groups using  $\pm 0.5$  points as the predefined  
522 smallest effect size of interest. Equivalence testing showed that the two experimental groups  
523 were statistically equivalent in narrative macrostructure post-intervention ( $p = .023, 90\% \text{ CI}$   
524  $[-0.42, 0.30]$ ), whereas neither experimental group was equivalent to the control group (control  
525 vs. narrative-based:  $p = .302, 90\% \text{ CI} [-0.76, 0.01]$ ; control vs. multimodal narrative-based:  $p$   
526  $= .402, 90\% \text{ CI} [-0.85, -0.03]$ ).

527 **[FIGURE 2 HERE]**

528 **Figure 2.** Mean narrative macrostructure scores broken down by Test (pre-intervention and  
529 post-intervention) and Group (control, narrative-based and multimodal narrative-based).

530 Asterisks represent significant differences: \*  $p \leq .05$ .

531

532

**Table 3.** Statistical results from the LME models for narrative macrostructure

<b>Fixed effects</b>	<b>df</b>	<b><math>\chi^2</math></b>	<b><i>p</i></b>
Test	1	1.06	.304
Group	2	4.46	.107
Test $\times$ Group	2	7.48	.024

533

534 **Narrative microstructure**535 *Total Number of Words (TNW)*

536 The model evaluating TNW reported a significant Test  $\times$  Group interaction. Post-hoc  
 537 comparisons showed first that while the narrative-based group had significantly more words at  
 538 post-intervention than at pre-intervention ( $d = .83, p < .001$ ), the control and multimodal  
 539 narrative-based groups did not. It also showed that at post-intervention, only the narrative-  
 540 based group outperformed the control group ( $d = -.90, p = .014$ ) (see Figure 3 and Table 4).

541 To further interpret the non-significant pre- to post-intervention differences in TNW in the  
 542 control and multimodal narrative-based groups, we conducted equivalence tests using the  
 543 conventional 80–125% ratio bounds as the smallest effect size of interest for word-count  
 544 measures. Both equivalence tests were significant (control:  $p = .029$ , ratio = .92, 90% CI [0.97,  
 545 1.23]; multimodal narrative-based:  $p = .002$ , ratio = .99, 90% CI [0.90, 1.14]), indicating that  
 546 pre–post changes in TNW were smaller than the predefined equivalence bounds and thus  
 547 practically negligible. At post-intervention, group comparisons were evaluated with the same  
 548 bounds of 0.80–1.25 (on the ratio scale). All equivalence tests were found to be non-significant  
 549 (control vs. narrative-based:  $p = .678$ , ratio = .77, 90% CI [0.66, 0.89]; control vs. multimodal  
 550 narrative-based:  $p = .262$ , ratio = .85, 90% CI [.72, 1.01]; narrative-based vs. multimodal  
 551 narrative-based:  $p = .091$ , ratio = 1.11, 90% CI [.96, 1.23]), which indicated that statistical  
 552 equivalence between groups post-intervention could not be established.

553

**[FIGURE 3 HERE]**

554 **Figure 3.** Mean microstructure values (Total Number of Words; TNW) broken down by Test  
 555 (pre-intervention and post-intervention) and Group (control, narrative-based and multimodal  
 556 narrative-based). Asterisks represent significant differences: \*  $p \leq .05$ , and \*\*\*  $p \leq .001$ .

557 *Total Number of Different Words (TNDW)*

558 Regarding the TNDW measure, only the two-way interaction was found to be significant. This  
559 suggested, on the one hand, that only the narrative-based group improved from pre- to post-  
560 intervention ( $d = .82, p < .001$ ) and, on the other, that the narrative-based group had  
561 significantly higher scores than the control group post-intervention ( $d = -.93, p = .016$ ). For  
562 the full statistical results report, see Table 4 (see also Figure 4).

563 To further examine the non-significant results comparing scores pre- to post-intervention, we  
564 conducted equivalence tests for TNDW, using the standard 80–125% bounds. The pre–post  
565 comparisons were significant for both the control group ( $p = .002$ , ratio = 0.94, 90% CI [0.97,  
566 1.16]) and the multimodal narrative-based group ( $p < .001$ , ratio = 1.00, 90% CI [0.93, 1.08]).  
567 This indicated that observed changes were smaller than the predefined equivalence threshold  
568 and providing evidence for the absence of meaningful change over time. The equivalence tests  
569 comparing groups at post-intervention were found to be significant, for the control vs.  
570 multimodal narrative-based comparison ( $p = .018$ , ratio = .92, 90% CI [0.83, 1.03]) and the  
571 narrative-based vs. multimodal narrative-based comparison ( $p = .017$ , ratio = 1.09, 90% CI  
572 [0.99, 1.21]), indicating that differences between these groups were smaller than the  
573 equivalence bound and supporting the conclusion that no practically meaningful group  
574 differences existed. The equivalence test for the comparison between the control and narrative-  
575 based groups was found to be non-significant ( $p = .191$ , ratio = .85, 90% CI [0.76, 0.94]), which  
576 indicated that there was no statistical equivalence between these two groups.

577 **[FIGURE 4 HERE]**

578 **Figure 4.** Mean microstructure values (Total Number of Different Words; TNDW) broken  
579 down by Test (pre-intervention and post-intervention) and Group (control, narrative-based  
580 and multimodal narrative-based). Asterisks represent significant differences: \*  $p \leq .05$ , and

581 \*\*\*  $p \leq .001$ .

582

583

**Table 4.** Statistical results from the LME models for narrative microstructure

Measure	Fixed effects	df	$\chi^2$	<i>p</i>
TNW	Test	1	1.98	.16
	Group	2	2.32	.314
	Test × Group	2	14.26	< .001
TNDW	Test	1	1.52	.217
	Group	2	1.25	.537
	Test × Group	2	14.99	< .001

584

585 **Narrative perspective-taking**

586 Regarding the narrative perspective-taking skills, only a main effect of Test was found to be  
 587 significant, indicating that scores were overall better pre-intervention than post-intervention,  
 588 regardless of group ( $d = .30, p = .025$ ) (see Figure 5 and Table 5).

589 The interaction between Test and Group was not found to be significant ( $p = .376$ ). For this,  
 590 we ran equivalence tests using  $\pm 0.5$  as the smallest meaningful difference to further interpret  
 591 the non-significant effects. The tests comparing the changes between pre- and post-intervention  
 592 were found to be significant for the three groups (control:  $p < .001$ , 90% CI  $[-0.02, 0.22]$ ;  
 593 narrative-based:  $p < .001$ , 90% CI  $[-0.10, 0.12]$ ; multimodal narrative-based:  $p < .001$ , 90% CI  
 594  $[-0.01, 0.25]$ ), which indicated that any observed changes were smaller than the predefined  
 595 meaningful difference. The equivalence tests for group comparisons at post-intervention did  
 596 not reach statistical equivalence (control vs. narrative-based:  $p = .183$ , 90% CI  $[-0.16, 0.02]$ ;  
 597 control vs. multimodal narrative-based:  $p = .304$ , 90% CI  $[-0.23, 0.01]$ ; narrative-based vs.  
 598 multimodal narrative-based:  $p = .056$ , 90% CI  $[-0.18, 0.09]$ ). Together these results indicate  
 599 that although the groups did not differ significantly at post-intervention, their scores could not  
 600 be considered statistically equivalent within the  $\pm 0.5$  bounds.

601

602

603

[FIGURE 5 HERE]

604

**Figure 5.** Mean narrative perspective-taking scores broken down by Test (pre-intervention and post-intervention) and Group (control, narrative-based and multimodal narrative-based).

605

606

**Table 5.** Statistical results from the LME models for narrative perspective-taking

<b>Fixed effects</b>	<b>df</b>	<b><math>\chi^2</math></b>	<b><i>p</i></b>
Test	1	17.81	< .001
Group	2	15.07	< .001
Test × Group	2	11.25	.004
Test	1	4.84	.028
Group	2	5.33	.07
Test × Group	2	1.96	.376

607

### 608 **Exploratory assessment of potential differences between the experimental groups**

609 Overall, the above-mentioned LME models indicated, overall, (1) that the narrative-based  
 610 group improved from pre- to post-intervention in narrative macrostructure and microstructure  
 611 and (2) that post-intervention their TNW and TNDW were significantly larger than those of  
 612 the control group. No significant effects were observed for the multimodal narrative-based  
 613 group. These findings were not in line with our predictions that the multimodal narrative-based  
 614 group would show greater improvements than those who receive the same narrative-based  
 615 intervention without the controlled multimodal component. Therefore, to further understand  
 616 why our hypothesis was not supported and what was making the multimodal narrative-based  
 617 group not improve, we conducted a complementary assessment of the implementation of the  
 618 intervention.

619 A research assistant who was blinded to the intervention group watched 18 intervention  
 620 sessions (i.e., 33% of the sessions) and evaluated children's and teachers' behavior throughout  
 621 the sessions in terms of (a) the teachers' effort to maintain children's attention and  
 622 concentration, and (b) the number of times they had to ask children to stay silent, pay attention

623 to the activity, or sit down. On a Likert scale of 1 to 7 (1 = low effort, no need to intervene to  
624 ask children to pay attention and 7 = high effort, constant intervention to ask children to pay  
625 attention), the results showed the teachers in the multimodal narrative-based group had to  
626 intervene way more often ( $M = 3.67$ ,  $SD = 1.87$ ) than those in the narrative-based group ( $M =$   
627  $1.67$ ,  $SD = 0.50$ ). Although there was no time difference between the two groups in relation to  
628 the duration of the sessions (approximately 25 to 30 minutes), it was also observed that teachers  
629 in the multimodal narrative-based group had to ask for silence, for them to sit down or pay  
630 attention an average of 13.67 times per session ( $SD = 10.61$ ), while those in the narrative-based  
631 group only an average of 1.44 times per session ( $SD = 1.51$ ). These results showed that the  
632 classroom dynamics of the multimodal narrative-based group was more disruptive than the one  
633 of the narrative-based group.

#### 634 **Narrative skills of children with language and communication needs**

635 Apart from the general narrative assessment for all children, we descriptively assessed whether  
636 children with language and communication needs were also able to show improvements by just  
637 receiving the tier 1 instruction. From the whole sample, a total of nine children were identified  
638 by the schools as having language and communication difficulties (control:  $N = 3$ ; narrative-  
639 based:  $N = 2$ ; multimodal narrative-based:  $N = 4$ ).

640 Due to the small sample size, no inferential statistics were conducted. However, from the  
641 descriptive statistics, we observed that whereas those in the control group showed no  
642 improvement, those in the experimental groups showed gains in narrative macrostructure and  
643 microstructure (*macrostructure*:  $M_{gain} = 0.5-1$  points; *TNW*:  $M_{gain} = 15.5-18.5$  words; *TNDW*:  
644  $M_{gain} = 5$  words) (see Table 6 for the descriptive statistics). A general tendency for children  
645 with language and communication needs to show greater narrative gains than children with  
646 typical development was also observed (see Table 7 for a summary of the average gains).

647 **Table 6.** Descriptives for pre- and post-intervention narrative assessments and the average gains between the pre- and post-intervention  
 648 assessments of children with language and communication needs, broken down by group and narrative measure

	Value	Control			Narrative-based			Multimodal narrative-based		
		PRE	POST	GAIN	PRE	POST	GAIN	PRE	POST	GAIN
<b>MACRO.</b>	M (SD)	2.17 (2.25)	1.83 (1.53)	-0.34	2.25 (3.18)	3.25 (1.77)	1	3.63 (0.63)	4.13 (1.80)	0.5
	Observed range	0–4.5	0.5–3.5	–	0–4.5	2–4.5	–	3–4.5	1.5–5.5	–
<b>MICRO.: TNW</b>	M (SD)	35.00 (13.44)	36.75 (23.69)	1.75	48.00 (0)	66.50 (0)	18.5	31.00 (6.24)	46.50 (7.76)	15.5
	Observed range	26–45	20–54	–	48–48	67–67	–	26–38	39–54	–
<b>MICRO.: TNDW</b>	M (SD)	23.75 (3.18)	23.00 (9.90)	-0.75	22.00 (0)	29.00 (0)	5	22.33 (6.45)	27.33 (4.51)	5
	Observed range	21–26	16–30	–	22–22	29–29	–	17–29.5	23–32	–
<b>PERS.</b>	M (SD)	0 (0)	0(0)	0	0.25 (0.35)	0 (0)	0	0.38 (0.48)	0 (0)	-0.38
	Observed range	0	0	–	0–0.5	0	–	0–1	0	–

649 *Note 1.* MACRO. = macrostructure; MICRO. = microstructure; TNW = total number of words; TNDW = total number of different words; PERS.  
 650 = perspective-taking; PRE = pre-intervention assessment; POST = post-intervention assessment

651 *Note 2.* The possible range for narrative macrostructure and narrative perspective-taking was 0–6. The possible range for the narrative  
652 microstructure measures was 0–NA, with NA indicating that no theoretical maximum value is defined for this measure.

653 *Note 3.* The score corresponds to the average score of the two narrations.

EN REVISIÓ

654 **Table 7.** Average gains between the the pre- and post-intervention assessments of children  
 655 with typical development and children with language and communication needs, broken  
 656 down by group and narrative measure

	<b>Control</b>		<b>Narrative-based</b>		<b>Multimodal narrative-based</b>	
	<b>TD</b>	<b>LCN</b>	<b>TD</b>	<b>LCN</b>	<b>TD</b>	<b>LCN</b>
<b>MACRO.</b>	-0.2	-0.34	0.34	1	0.08	0.5
<b>MICRO.: TNW</b>	-4.46	1.75	8.51	18.5	-0.95	15.5
<b>MICRO.: TNDW</b>	-2.13	-0.75	3.84	5	-0.33	5
<b>PERS.</b>	-0.13	0	0	0	-0.09	-0.38

657 *Note.* MACRO. = macrostructure; MICRO. = microstructure; TNW = total number of words;  
 658 TNDW = total number of different words; PERS. = perspective-taking; TD = children with  
 659 typical development; LCN = children with language and communication needs

660

## DISCUSSION

661 The present study aimed to assess the effectiveness of the [program], particularly its core  
662 multimodal component, in fostering preschool children's oral narrative skills. To do this, we  
663 implemented the tier 1 of the [program] with 36 children attending the last year of preschool  
664 in three schools in a city in the metropolitan area of Barcelona. The performance of these 36  
665 children was compared to the performance of a group of 41 children receiving the [program]  
666 without the multimodal component and to the performance of a treatment-as-usual control  
667 group with 38 children. Our results revealed that only the children in the narrative-based  
668 significantly improved their oral narrative macrostructure, and microstructure skills, and that  
669 they both outperformed the control group not receiving the intervention in the microstructure  
670 measures. These findings thus reject our hypothesis that the children receiving the intervention  
671 with the specific multimodal component would show greater improvements than those in the  
672 narrative-based group.

673 As just mentioned, the narrative-based group showed significant pre–post improvements in the  
674 narrative macrostructure and microstructure measures. We believe that these findings might  
675 contribute to the existing evidence on group-based narrative intervention research reporting its  
676 potential to boost young children's oral narrative skills (e.g., Daelman et al., 2024; Gillam,  
677 Olszewski et al., 2014; Nelson et al., 2022; Pinto et al., 2019; Spencer et al., 2018; West et al.,  
678 2021; for reviews, see Pico et al., 2021). Several factors may account for the significant  
679 improvements observed in the narrative-based group. On the one hand, the intervention was  
680 designed in a co-creation fashion with a group of more than 90 preschool teachers and speech-  
681 language therapists, such that it was tailored to the Catalan context, integrating professionals'  
682 usual practices and needs. On the other hand, the [program] aligns with the narrative  
683 intervention principles postulated by Spencer and Petersen (2020), as well as with the strategies  
684 commonly used in narrative intervention studies described in the recent reviews by Favot et al.  
685 (2021) and Pico et al. (2021). These principles and strategies include specific instruction on the  
686 macrostructure (i.e., story elements), use of visual support such as pictures or icons, and verbal  
687 prompting. The [program] was designed considering previous literature, and therefore includes  
688 strategies such as training story elements through a verbal question-and-answer sequence,  
689 providing oral feedback, and using audiovisuals such as video cartoons, icons, and animated  
690 pictures.

691 The current study not only assessed narrative macrostructure and microstructure, but also  
692 narrative perspective-taking, understood as children's ability to talk about the emotions and  
693 perspectives of the characters. We observed that the scores for this measure were quite low,  
694 ranging from 0 to 2 when they could go up to 6, and that there were no significant improvements  
695 in this measure after the intervention. We believe that even though children are able to  
696 recognize and verbalize emotions at the age of three (see Widen & Nelson, 2022 for a review),  
697 being able to include them when narrating a story is a more challenging task that is not acquired  
698 until later in development, around the age of eight to nine (e.g., Veneziano et al., 2020).  
699 Building a narration is complex because it involves remembering story events, expressing them  
700 in order, and using appropriate vocabulary. Therefore, if on top of that, children also need to  
701 remember the emotions and perspectives of the characters, which are less salient aspects of the  
702 story compared to the macrostructural elements, this makes it even a more complex task. For  
703 this reason, we believe that preschool children might find it too challenging to talk about the  
704 story plot and the characters' emotions and perspectives at the same time, and for this reason,  
705 they focus mainly on describing what happened (i.e., macrostructure).

706 Complementarily, the present study also sought to assess the effectiveness of the multimodal  
707 component within the [program]. Based on the multimodal enrichment paradigm (Mathias &  
708 von Kriegstein, 2023) and on the great amount of research suggesting the causal role of  
709 multimodality on children's language development (for reviews, see Dargue et al., 2019;  
710 Goldin-Meadow, 2014; Hostetter, 2011; Rohlfing, 2017; Vilà-Giménez & Prieto, 2021), we  
711 expected significant improvements of the group receiving the [program] with the multimodal  
712 component. Nevertheless, our study showed no significant narrative improvements in this  
713 group.

714 There are two possible explanations for our findings on the lack of observed effects of the  
715 multimodal narrative-based group, which can be related to the naturalistic nature of the study.  
716 First, it should be taken into account that many studies assessing the causal role of  
717 multimodality (particularly gesture) in children's language learning were lab-based studies in  
718 which participants were trained individually in a very controlled environment where they either  
719 saw an audiovisual stimulus of someone gesturing (i.e., like the videorecordings of the  
720 storyteller within the [program]) or someone standing still. However, as also reported recently  
721 by Bernstein and colleagues (2024), this isolated testing of gesture effects can be difficult in a  
722 naturalistic environment, such as a classroom, provided that interventions like the one  
723 implemented in this study or the one implemented by Bernstein et al. (2024) are naturalistic

724 interventions that are integrated into the regular teaching curriculum in the school setting. Also,  
725 related to the naturalistic implementation of the intervention, there is a second factor that could  
726 explain our findings. We thought that the specific way preschool teachers implemented the  
727 intervention might have played a key role and for this, a complementary assessment of the  
728 implementation of the intervention was conducted. These assessments showed that the  
729 classroom dynamics of the multimodal narrative-based group, which required children to  
730 repeatedly stand up, enact the stories, and sit down again, may have been too intense and thus,  
731 may have disrupted children's sustained attention, potentially limiting opportunities for  
732 narrative learning, whereas children receiving the narrative-based intervention might have been  
733 more focused on the activity as they were sitting most of the time. Bernstein et al. (2024) also  
734 acknowledged the need to observe the behavior by children and the particular implementation  
735 by the teacher to see if their behavior can be a potential explanation to their null findings, which  
736 is consistent with the complementary assessment of intervention implementation conducted in  
737 the present study. Therefore, our findings suggest that multimodality may be used in classroom  
738 settings with preschoolers when it is implemented in a manner that does not interfere with the  
739 dynamics of the teacher–group interaction, such as in the narrative-based group, where children  
740 are not asked to constantly stand up and sit down to enact the story elements. Another aspect  
741 that should be taken into account is that, although the intervention was co-created with  
742 preschool teachers and speech-language therapists who considered it feasible and appropriate  
743 for the context of implementation, children had limited experience with language-based and  
744 multimodal interactive activities in the classroom. In some contexts (like ours), children are  
745 usually sitting throughout the sessions and are not used to engaging in more active activities  
746 such as multimodal story enactment.

747 The results of the present study complement and expand on the results of previous  
748 investigations. As mentioned above, very few studies have systematically integrated and tested  
749 the effects of the multimodality component within a classroom-based narrative intervention,  
750 and those that have done so have reported mixed findings. In a classroom environment, existing  
751 studies have reported null results, similar to those reported in the present study with a lack of  
752 generalization effects. For instance, Pronina and colleagues implemented an intervention  
753 focusing on embodying mental states through facial expressions in a preschool setting and  
754 assessed how the intervention could boost pragmatic and socioemotional measures. They found  
755 that while the intervention was effective in improving children's expressive pragmatic skills,  
756 there was no difference between children receiving the embodied intervention and those

757 receiving the intervention without embodiment, specifically for emotion understanding and  
758 mental state verb comprehension, which were directly trained during the intervention.  
759 Similarly, Bernstein et al. (2024) compared children's narrative recall after receiving a  
760 storybook-reading intervention, with or without drama-based activities involving the teacher  
761 and children standing up and acting out as the main characters experiencing the main actions  
762 and events of the story plot. The authors found no significant differences between the two  
763 groups after the intervention in their ability to narrate a story. These two studies incorporated  
764 multimodality in a similar fashion as in the present study: involving teachers enacting the main  
765 elements of stories and involving children in also enacting the stories. For this, we believe that  
766 in our study, the fact of asking children to stand up and involve their whole body in the  
767 enactment might have also disrupted their concentration on the target skill. These findings seem  
768 to point out that on the one hand, it is more challenging to isolate the beneficial effects of  
769 multimodality when implemented in a more naturalistic context (such as a real classroom  
770 setting) and that in these settings it may be beneficial to provide professionals with additional  
771 guidance on implementing multimodality in a more structured and less intense manner so that  
772 it does not disrupt the communicative interaction between the teacher and the children.

773 This intervention study has certain limitations regarding its methodology and classroom  
774 implementation. First, regarding the assessments, post-intervention assessments were  
775 conducted at the end of the intervention period, which meant that the children were evaluated  
776 approximately after one to two weeks. Therefore, the results of this study correspond to the  
777 immediate effects of the intervention. A delayed post-intervention assessment would have been  
778 useful for determining whether the intervention effects remained over time. In fact, this delayed  
779 assessment was planned before the start of data collection but could not be implemented  
780 because of school schedule restrictions which resulted in the dropout of one school. Another  
781 limitation of this study is related to the narrative macrostructure coding scheme used: although  
782 story complexity was controlled for, the first story featured only one character and the second  
783 featured two characters. The coding scheme used in this study did not distinguish whether  
784 children introduced one or two characters. Future studies should use coding criteria that are  
785 more fine-grained and can reflect these differences. Also, the effects for narrative perspective-  
786 taking were small. Given the overall low scores across all groups, care should be taken not to  
787 overinterpret the control group's post-intervention score of zero as evidence of a complete  
788 absence of this skill. This may reflect a general floor effect or limited sensitivity of the current  
789 measurement approach. It would also be interesting to conduct more fine-grained analyses of

790 children's narrative perspective-taking to assess which specific emotional or mental terms  
791 children are able to identify and use more frequently. Following up on the coding procedures,  
792 it should be acknowledged that experimenters were not blind to the time of testing (i.e., pre- or  
793 post-intervention). Although this awareness could have introduced bias, this risk was minimal  
794 because of the large sample size and the time elapsed between the pre and post sessions (one  
795 to two months), making it highly unlikely that coders recalled individual children's earlier  
796 performances. Nevertheless, future studies could address this limitation by ensuring that coding  
797 is performed by independent raters who are blind to testing time points.

798 Second, with regard to the classroom implementation, this study tested the implementation of  
799 a tier 1 intervention for all children in the classroom. Although there were nine children with  
800 language and communication difficulties within the sample, a full multi-tiered intervention  
801 could not be implemented because of the difficulty in coordinating both classroom intervention  
802 and small-group or individual interventions. Therefore, these children only received the tier 1  
803 instruction without any extra support. Despite this, as mentioned in the Results section, these  
804 children still showed improvements. In addition, we acknowledge that the intervention period  
805 was relatively short (only nine sessions over three weeks). A longer intervention might provide  
806 more opportunities for children to familiarize themselves with the structure and activities of  
807 the intervention, and for teachers to integrate the approach naturally within their curriculum.  
808 Particularly for the multimodal strategies, we believe that children may need more prior  
809 experience using their bodies to enact the story, especially in contexts where such methods are  
810 not typically used. Including drama-based exercises as a preparatory step before introducing a  
811 multimodal narrative intervention could help children engage more fully and naturally with the  
812 intervention. For all this, we think that future studies should test whether a longer intervention  
813 could help obtain stronger effects.

814 Several considerations should be taken into account related to the multimodal component of  
815 the intervention and the complementary assessment of its implementation. While involving  
816 professionals in the design and implementation of the intervention from a participatory  
817 approximation to research added ecological validity to the study, the way in which the  
818 multimodal component was implemented may have limited its effectiveness. This likely  
819 reflects challenges in applying these multimodal strategies in practice, rather than any  
820 shortcoming on the part of the professionals. We believe it would have been interesting to first  
821 run a pre-study with a single trained interventionist delivering all groups of the intervention to  
822 better control implementation and identify any potential problematic issues before the full-

823 scale implementation. Additionally, we would like to acknowledge that the complementary  
824 assessment of intervention implementation was conducted retrospectively to explore  
825 differences across the experimental groups, rather than integrated into the original study design.  
826 Nonetheless, this assessment offered useful insights and directions for improving the  
827 integration of multimodality in naturalistic intervention settings in the future. For example, to  
828 reduce classroom disruptions and better support children's attention to the language-based  
829 interaction, it may be helpful to allocate a dedicated time slot to enact the full story, or to enact  
830 actions and emotions while seated. Importantly, while the current study is not a definitive test  
831 of the isolated effects of multimodality, it does contribute valuable evidence on the general  
832 impact of narrative interventions in authentic educational settings and offers guidance for  
833 improving the structured implementation of multimodal components in future research.

834 The findings of this study have implications for classroom intervention practice and research.  
835 It was shown that a short 9-session narrative intervention program can be effective in fostering  
836 preschool children's oral narrative macrostructure and microstructure skills. Regarding the  
837 particular implementation of multimodality in classroom environments, educators should pay  
838 attention to the fact that multimodality needs to be well integrated into narrative interventions  
839 in a way that it complements (and does not disrupt) communicative interactions, as these  
840 integration mechanisms can have impactful consequences for children's learning. Thus, our  
841 findings offer insights into consideration for implementing multimodality within narrative  
842 intervention programs in preschool education.

- 844 Babayiğit, S., Roulstone, S. & Wren, Y. (2021), Linguistic comprehension and narrative skills  
845 predict reading ability: A 9-year longitudinal study. *British Journal of Educational*  
846 *Psychology*, 91, 148–168. <https://doi.org/10.1111/bjep.12353>
- 847 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models  
848 Using lme4. *Journal of Statistical Software*, 67(1), 1–48.  
849 <https://doi.org/10.18637/jss.v067.i01>
- 850 Bernstein, K. A., van Huisstede, L., Marley, S. C., Gao, Y. B., Pierce-Rivera, M., Ippolito, E.,  
851 Restrepo, M. A., Millinger, J., Brantley, K., & Gantwerker, J. (2024). Gesture like a kitten  
852 and you won't forget your tale: Drama-based, embodied story time supports preschoolers'  
853 narrative skills. *Early Childhood Research Quarterly*, 66, 178–190.  
854 <https://doi.org/10.1016/j.ecresq.2023.10.004>
- 855 blinded (2025a)
- 856 blinded (2025b)
- 857 Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.
- 858 Corbett, B. A., Key, A. P., Qualls, L., Fecteau, S., Newsom, C., Coke, C., & Yoder, P. (2016).  
859 Improvement in Social Competence Using a Randomized Trial of a Theatre Intervention for  
860 Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*,  
861 46(2), 658–672. <https://doi.org/10.1007/s10803-015-2600-9>
- 862 Daelman, J., Alighieri, C., Van Lierde, K., Simon, E., Altinkamis, F., Baudonck, N., &  
863 D'haeseleer, E. (2023). Effect of a Group-Based Narrative Language Training in Typically  
864 Developing Turkish-Dutch Children. *Folia Phoniatica et Logopaedica*, 76(2), 192–205.  
865 <https://doi.org/10.1159/000533620>
- 866 Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-  
867 analysis into the effects of gesture on comprehension. *Psychological Bulletin*, 145(8), 765–  
868 784. <https://doi.org/10.1037/bul0000202>
- 869 Demir, Ö. E., Fisher, J. A., Goldin-Meadow, S., & Levine, S. C. (2014). Narrative processing  
870 in typically developing children and children with early unilateral brain injury: Seeing

- 871 gesture matters. *Developmental Psychology*, 50(3), 815–828.  
872 <https://doi.org/10.1037/a0034322>
- 873 Demir, Ö. E., Levine, S. C., & Goldin-Meadow, S. (2015). A tale of two hands: Children's  
874 early gesture use in narrative production predicts later narrative structure in speech. *Journal*  
875 *of Child Language*, 42(3), 662–681. <https://doi.org/10.1017/S0305000914000415>
- 876 Dickinson, D. K., & McCabe, A. (2001). Bringing it all together: The multiple origins, skills,  
877 and environmental supports of early literacy. *Learning Disabilities Research & Practice*,  
878 16(4), 186–202.
- 879 Dodd, J. L., Ocampo, A., & Kennedy, K. S. (2011). Perspective taking through narratives: An  
880 intervention for students with ASD. *Communication Disorders Quarterly*, 33(1), 23–33.  
881 <https://doi.org/10.1177/1525740110395014>
- 882 Donolato, E., Toffalini, E., Rogde, K., Nordahl-Hansen, A., Lervåg, A., Norbury, C., & Melby-  
883 Lervåg, M. (2023). Oral language interventions can improve language outcomes in children  
884 with neurodevelopmental disorders: A systematic review and meta-analysis. *Campbell*  
885 *Systematic Reviews*, 19(4), Article e1368. <https://doi.org/10.1002/cl2.1368>
- 886 Douglas, K. (2019). *Improving Narrative and Expository Language: A Comparison of*  
887 *Narrative Intervention to Shared Storybook Reading*. (Publication No. 28103454) [Master  
888 dissertation, Wilmington University]. ProQuest Dissertations & Theses Global.
- 889 Esteve-Gibert, N., & Prieto, P. (2018). *The Development of Prosody in First Language*  
890 *Acquisition*. John Benjamins.
- 891 Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using  
892 G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*,  
893 41, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- 894 Frydman, J. S., & Mayor, C. (2024). A Scoping Review on the Use and Potential of School-  
895 Based Drama Therapy to Enhance Socio-emotional Skills in Early Childhood. *Early*  
896 *Childhood Education Journal*, 52(4), 669–680. [https://doi.org/10.1007/s10643-023-01471-](https://doi.org/10.1007/s10643-023-01471-1)  
897 [1](https://doi.org/10.1007/s10643-023-01471-1)

- 898 Gillam, S. L., Gillam, R. B., & Laing, C. E. (2014). *Supporting knowledge in language and*  
899 *literacy*. Logan, UT: Utah State University.
- 900 Gillam, S. L., Olszewski, A., Fargo, J., & Gillam, R. B. (2014). Classroom-Based Narrative  
901 and Vocabulary Instruction: Results of an Early-Stage, Nonrandomized Comparison Study.  
902 *Language, Speech, and Hearing Services in Schools, 45*(3), 204–219.  
903 [https://doi.org/10.1044/2014\\_LSHSS-13-0008](https://doi.org/10.1044/2014_LSHSS-13-0008)
- 904 Goldin-Meadow, S. (2014). Widening the lens: What the manual modality reveals about  
905 language, learning and cognition. *Philosophical Transactions of the Royal Society B:*  
906 *Biological Sciences, 369*(1651). <https://doi.org/10.1098/rstb.2013.0295>
- 907 Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral Discourse in the Preschool  
908 Years and Later Literacy Skills. *First Language, 24*(2), 123–147.  
909 <https://doi.org/10.1177/0142723704042369>
- 910 Guilbert, D., Sweller, N., & Van Bergen, P. (2021). Emotion and gesture effects on narrative  
911 recall in young children and adults. *Applied Cognitive Psychology, 35*(4), 873–889.  
912 <https://doi.org/10.1002/acp.3815>
- 913 Henry, A. R., & Solari, E. J. (2020). Targeting Oral Language and Listening Comprehension  
914 Development for Students with Autism Spectrum Disorder: A School-Based Pilot Study.  
915 *Journal of Autism and Developmental Disorders, 50*(10), 3763–3776.  
916 <https://doi.org/10.1007/s10803-020-04434-2>
- 917 Hettiarachchi, S. (2022). The effectiveness of the adapted Box Clever language intervention  
918 programme in the development of vocabulary and narrative skills of deaf and hard of hearing  
919 children. *Deafness & Education International, 24*(1), 65–91.  
920 <https://doi.org/10.1080/14643154.2020.1721158>
- 921 Hostetter, A. B. (2011). When Do Gestures Communicate? A Meta-Analysis. *Psychological*  
922 *Bulletin, 137*(2), 297–315. <https://doi.org/10.1037/a0022128>
- 923 Ionescu, T., & Ilie, A. (2018). Language learning in preschool children: An embodied learning  
924 account. *Early Child Development and Care, 188*(1), 4–15.  
925 <https://doi.org/10.1080/03004430.2016.1189419>
- 926 Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test*. Pearson.

- 927 Khan, K., Nelson, K., & Whyte, E. (2014). Children choose their own stories: The impact of  
928 choice on children's learning of new narrative skills. *Journal of Child Language*, 41(4),  
929 949–962. <https://doi.org/10.1017/S0305000913000160>
- 930 Kirk, E., Donnelly, S., Furman, R., Warmington, M., Glanville, J., & Eggleston, A. (2022).  
931 The relationship between infant pointing and language development: A meta-analytic  
932 review. *Developmental Review*, 64, Article 101023.  
933 <https://doi.org/10.1016/j.dr.2022.101023>
- 934 Lakens, D. (2017). *TOSTER: Two one-sided tests (TOST) equivalence testing*. R package  
935 version 0.3.4. <https://CRAN.R-project.org/package=TOSTER>
- 936 Lee, B. K., Patall, E. A., Cawthon, S. W., & Steingut, R. R. (2015). The Effect of Drama-Based  
937 Pedagogy on PreK–16 Outcomes: A Meta-Analysis of Research From 1985 to 2012. *Review*  
938 *of Educational Research*, 85(1), 3–49. <https://doi.org/10.3102/0034654314540477>
- 939 Li, J. (2024). *The use of referential iconic character viewpoint gestures as an indicator of child-*  
940 *directed narrative speech* [Master thesis]. Universitat Pompeu Fabra.
- 941 Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., Makowski, D. (2021). performance: An  
942 R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open*  
943 *Source Software*, 6(60), Article 3139. <https://doi.org/10.21105/joss.03139>
- 944 Macoun, A., & Sweller, N. (2016). Listening and watching: The effects of observing gesture  
945 on preschoolers' narrative comprehension. *Cognitive Development*, 40, 68–81.  
946 <https://doi.org/10.1016/j.cogdev.2016.08.005>
- 947 MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition.  
948 Lawrence Erlbaum Associates.
- 949 Mathias, B., & von Kriegstein, K. (2023). Enriched learning: Behavior, brain, and computation.  
950 *Trends in Cognitive Sciences*, 27(1), 81–97. <https://doi.org/10.1016/j.tics.2022.10.007>
- 951 McHugh M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–  
952 282.
- 953 McKern, N., Dargue, N., Sweller, N., Sekine, K., & Austin, E. (2021). Lending a hand to  
954 storytelling: Gesture's effects on narrative comprehension moderated by task difficulty and

- 955 cognitive ability. *Quarterly Journal of Experimental Psychology*, 74(10), 1791–1805.  
956 <https://doi.org/10.1177/17470218211024913>
- 957 Murachver, T., Pipe, M.-E., Gordon, R., Owens, J. L., & Fivush, R. (1996). Do, Show, and  
958 Tell: Children’s Event Memories Acquired Through Direct Experience, Observation, and  
959 Stories. *Child Development*, 67(6), 3029–3044. <https://doi.org/10.2307/1131765>
- 960 Nelson, K. E., & Khan, K. S. (2019). A dynamic systems account incorporating eight narrative  
961 developmental stages: Chapter 8. New frontiers in facilitating narrative skills in children  
962 and adolescents. In E. Veneziano & A. Nicolopoulou (Eds.), *Narrative, Literacy and Other*  
963 *Skills: Studies in intervention* (pp. 173–199). John Benjamins Publishing Company.  
964 <https://doi.org/10.1075/sin.25.09nel>
- 965 Nelson, B. S., Petersen, D. B., & Rai, A. (2022). The effects of a multi-tiered system of  
966 language support on oral narrative language, writing, and reading comprehension in India.  
967 *Language and Education*, 36(1), 74–94. <https://doi.org/10.1080/09500782.2021.1898633>
- 968 Nicolopoulou, A., Schnabel, K., Ilgaz, H., Brockmeyer, C., & De Sá, A. B. (2015). Using a  
969 narrative- and play-based activity to promote low-income preschoolers’ oral language,  
970 emergent literacy, and social competence. *Early Childhood Research Quarterly*, 31, 147–  
971 162. <https://doi.org/10.1016/j.ecresq.2015.01.006>
- 972 Parrill, F., Lavanty, B., Bennett, A., Klco, A., & Demir-Lira, O. E. (2018). The relationship  
973 between character viewpoint gesture and narrative structure in children. *Language and*  
974 *Cognition*, 10(3), 408–434. <https://doi.org/10.1017/langcog.2018.9>
- 975 Perniss, P. (2018). Why we should study multimodal language. *Frontiers in Psychology*,  
976 9(JUN), 1–5. <https://doi.org/10.3389/fpsyg.2018.01109>
- 977 Petersen, D. B., Staskowski, M., Spencer, T. D., Foster, M. E., & Brough, M. P. (2022). The  
978 Effects of a Multi-tiered System of Language Support on Kindergarten Oral and Written  
979 Language: A Large-Scale Randomized Controlled Trial. *Language, Speech, and Hearing*  
980 *Services in Schools*, 53(1), 44–68. [https://doi.org/10.1044/2021\\_LSHSS-20-00162](https://doi.org/10.1044/2021_LSHSS-20-00162)
- 981 Pico, D. L., Hessling Prah, A., Biel, C. H., Peterson, A. K., Biel, E. J., Woods, C., & Contesse,  
982 V. A. (2021). Interventions Designed to Improve Narrative Language in School-Age  
983 Children: A Systematic Review With Meta-Analyses. *Language, Speech, and Hearing*  
984 *Services in Schools*, 52(4), 1109–1126. [https://doi.org/10.1044/2021\\_LSHSS-20-00160](https://doi.org/10.1044/2021_LSHSS-20-00160)

- 985 Pinto, G., Tarchi, C., & Bigozzi, L. (2019). Promoting narrative competence in kindergarten:  
986 An intervention study. *Early Childhood Research Quarterly*, 47, 20–29.  
987 <https://doi.org/10.1016/j.ecresq.2018.09.003>
- 988 Pronina, M., Hübscher, I., Holler, J., & Prieto, P. (2021). Interactional training interventions  
989 boost children’s expressive pragmatic abilities: Evidence from a novel multidimensional  
990 testing approach. *Cognitive Development*, 57, Article 101003.  
991 <https://doi.org/10.1016/j.cogdev.2020.101003>
- 992 Rohlfing, K. (2019). Learning language from the use of gestures. In J. S. Horst & J. V.  
993 Torkildsen (Eds.), *International Handbook of Language Acquisition* (pp. 213–233).  
994 Routledge.
- 995 Seccia, A., & Goldin-Meadow, S. (2024). Gestures can help children learn mathematics: How  
996 researchers can work with teachers to make gesture studies applicable to classrooms.  
997 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 379(1911), Article  
998 20230156. <https://doi.org/10.1098/rstb.2023.0156>
- 999 Spencer, T. D., & Petersen, D. B. (2020). Narrative Intervention: Principles to Practice.  
1000 *Language, Speech, and Hearing Services in Schools*, 51(4), 1081–1096.  
1001 [https://doi.org/10.1044/2020\\_LSHSS-20-00015](https://doi.org/10.1044/2020_LSHSS-20-00015)
- 1002 Spencer, T. D., & Petersen, D. B. (2018a). *Story Champs® 2.0*. Laramie, WY: Language  
1003 Dynamics Group. [https://www.languagedynamicsgroup.com/story-champs-2/story-](https://www.languagedynamicsgroup.com/story-champs-2/story-champs-about/)  
1004 [champs-about/](https://www.languagedynamicsgroup.com/story-champs-2/story-champs-about/)
- 1005 Spencer, T. D., & Petersen, D. B. (2018b). Narratives bridge oral and written language: An oral  
1006 language intervention study with writing outcomes. *Language, Speech, and Hearing*  
1007 *Services in Schools*, 49, 569-581. [https://doi.org/10.1044/2018\\_LSHSS-17-0030](https://doi.org/10.1044/2018_LSHSS-17-0030)
- 1008 Spencer, T. D., Weddle, S. A., Petersen, D. B., & Adams, J. A. (2018). Multi-tiered narrative  
1009 intervention for preschoolers: A Head Start implementation study. *NHSA Dialog*, 20(1), 1–  
1010 28.
- 1011 Suggate, S. P., Lenhart, J., Vaahtoranta, E., & Lenhard, W. (2021). Interactive elaborative  
1012 storytelling fosters vocabulary in preschoolers compared to repeated-reading and phonemic  
1013 awareness interventions. *Cognitive Development*, 57, Article 100996.  
1014 <https://doi.org/10.1016/j.cogdev.2020.100996>

- 1015 Veneziano, E., Le Normand, M.-T., Plumet, M.-H., & Elie-Deschamps, J. (2020). Promoting  
1016 narrative skills in 5- to 8-year-old French-speaking children: The effects of a short  
1017 conversational intervention. *First Language*, 40(3), 225–250.  
1018 <https://doi.org/10.1177/0142723720901614>
- 1019 Vilà-Giménez, I., Igualada, A., & Prieto, P. (2019). Observing storytellers who use rhythmic  
1020 beat gestures improves children’s narrative discourse performance. *Developmental*  
1021 *Psychology*, 55(2), 250–262. <https://doi.org/10.1037/dev0000604>
- 1022 Vilà-Giménez, I., & Prieto, P. (2021). The Value of Non-Referential Gestures: A Systematic  
1023 Review of Their Cognitive and Linguistic Effects in Children’s Language Development.  
1024 *Children*, 8(2), Article 148. <https://doi.org/10.3390/children8020148>
- 1025 Vilà-Giménez, I., & Prieto, P. (2020). Encouraging kids to beat: Children’s beat gesture  
1026 production boosts their narrative performance. *Developmental Science*, 23(6), 1–14.  
1027 <https://doi.org/10.1111/desc.12967>
- 1028 Walker, D., Sepulveda, S. J., Hoff, E., Rowe, M. L., Schwartz, I. S., Dale, P. S., ... & Bigelow,  
1029 K. M. (2020). Language intervention research in early childhood care and education: A  
1030 systematic survey of the literature. *Early Childhood Research Quarterly*, 50, 68–85.  
1031 <https://doi.org/10.1016/j.ecresq.2019.02.010>
- 1032 West, G., Snowling, M. J., Lervåg, A., Buchanan-Worster, E., Duta, M., Hall, A., McLachlan,  
1033 H., & Hulme, C. (2021). Early language screening and intervention can be delivered  
1034 successfully at scale: Evidence from a cluster randomized controlled trial. *Journal of Child*  
1035 *Psychology and Psychiatry, and Allied Disciplines*, 62(12), 1425–1434.  
1036 <https://doi.org/10.1111/jcpp.13415>
- 1037 Widen, S. C., & Nelson, N. L. (2022). Differentiation and Language Acquisition in Children’s  
1038 Understanding of Emotion. In D. Dukes, A. C. Samson & E. A. Walle (Eds.), *The Oxford*  
1039 *Handbook of Emotional Development* (pp. 174–187). Oxford Academic.
- 1040 Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language*  
1041 *Fundamentals—fifth edition (CELF-5)*. Pearson.
- 1042 Winter, B. (2020). *Statistics for Linguists: An Introduction Using R*. Routledge.
- 1043

1044 **DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN**  
1045 **THE MANUSCRIPT PREPARATION PROCESS**

1046 During the preparation of this work the authors used PaperPal in order to proofread the  
1047 manuscript and improve its readability. After using these tools, the authors reviewed and edited  
1048 the content as needed and take full responsibility for the content of the published article.

EN REVISIÓ