Revista Chilena de Fonoaudiología 24(1), 1-18, 2025
https://revfono.uchile.cl/
ISSN 0719-4692

Original Article

Narrative Development Assessment: Contributions from Research on Children with DLD and TD

María Luisa Silva [a, *], Nina Crespo-Allende [b], Jeannette Sepúlveda-Toro [b], Lorena Pérez-Abarca [b] and Carola Alvarado-Barra [c]

[a] *Centro Interdisciplinario de Investigación en Psicología y Matemática Experimental CIIPME-CONICET, Argentina.*
[b] *Pontificia Universidad Católica de Valparaíso, Chile.*
[c] *Facultad de Salud, Universidad Santo Tomás, Chile.*

ABSTRACT

Research on the increasing complexity of children's narratives has provided insight into how various linguistic resources are organized. However, to better understand this development, it is necessary to consider not only the relationship between discursive and linguistic factors but also to include cognitive aspects. This study aims to delve into the complexity involved in measuring child narrative development. To this end, two stages of narrative development are described and compared in two groups of monolingual Spanish-speaking Chilean children: one with Developmental Language Disorder (DLD) and another with Typical Development (TD). The analysis incorporates psycholinguistic indicators and the Adapted Index of Narrative Complexity (INC-A). Sixteen children with DLD and 23 with TD participated in the study. The assessment consisted of retelling tasks performed at ages 5 and 10. Results from the psycholinguistic indicators showed no significant group differences at either time point. No statistically significant differences were found in intergroup proportions in the INC-A; however, time had a considerable effect on score proportions. Unlike previous studies, we only found significant intergroup differences in the Plan sequence.

Keywords:
Language Development; Narrative; Oral Production; Developmental Language Disorder; Criterion-Based Assessment

Evaluación del desarrollo narrativo: Aportes de una investigación en niños con TDL y con DT

RESUMEN

El estudio de la complejización de la narrativa infantil ha permitido comprender cómo se organizan diferentes recursos. No obstante, para comprender mejor este desarrollo, además de atender a la relación entre factores discursivo-lingüísticos, también se deben considerar factores cognitivos. Este trabajo propone profundizar en la complejización de la medición del desarrollo narrativo infantil. Para ello, se describen y comparan dos momentos del desarrollo narrativo de dos grupos de niños chilenos, monolingües hablantes de español, a saber, uno con Trastorno del Desarrollo del Lenguaje (TDL), y otro con Desarrollo Típico (DT), mediante la utilización de indicadores de productividad psicolingüística y el Índice de Complejidad Narrativa Adaptado-ICN-A. Participaron en el estudio 16 niños con TDL y 23 con DT, evaluados en tareas de recontado a los 5 y a los 10 años. Los resultados de los indicadores psicolingüísticos no manifiestan diferencias significativas entre grupos en ninguno de los momentos de la toma. Respecto al ICN-A, no se registran diferencias estadísticamente significativas en las proporciones intergrupo; aunque se observa que el tiempo tiene incidencia significativa al considerar las proporciones de puntuaciones. A diferencia de previos trabajos, solo se halló diferencia significativa intergrupo en la secuencia Plan.

Palabras clave:
Desarrollo Lingüístico; Narración; Producción Oral; Trastorno del Desarrollo del Lenguaje; Evaluación por Criterios

**Corresponding Author: María Luisa Silva*
Email: mlsilva@conicet.gov.ar

Received: 05-09-2023
Accepted: 01-27-2025
Published: 06-03-2025

INTRODUCTION

Assessing narrative performance entails analyzing both linguistic and communicative resources in response to the demand of constructing a story. Discourse-level features, in turn, enable inferences about underlying cognitive processes and pragmatic appropriateness (Volkmar, 2020). Narration involves constructing plausible events within discourse, whereas retelling includes retrieving a sequence of events from memory to reformulate the source narrative. Given that narrating integrates the roles of speaker, cognitive agent, and social subject (Bruner, 1986), assessments must address the complexity of these dimensions.

The evaluation of narrative production requires considering a range of skills and knowledge, including linguistic, discursive, cognitive, social, and identity-related aspects (Crespo, Silva, et al., 2021; Crespo & Silva, 2019). The design of assessment instruments should reflect this complexity (McCabe & Rollins, 1994). Westerveld & Gillon (1999) outline several conditions for elicitation tasks aimed at narrative retelling: the narrative should be produced for a listener unfamiliar with the story (Masterson & Kamhi, 1991); the stimulus should be presented on two separate occasions (Gummersall & Strong, 1999); and the stimuli should vary in linguistic complexity, operationalized by the number of events included (Griffith et al., 1986). Additionally, the task should elicit comparable productions that can be analyzed using consistent criteria (Silva, 2008, 2010). The retelling task satisfies many of these requirements, as it uses controlled stimuli (Romero Contreras & Gómez Martínez, 2013), makes it possible to differentiate between linguistic and cognitive demands, and yields productions that are comparable across individuals (Vivas Vivas et al., 2021).

There are various approaches to analyzing data obtained from narrative tasks (Mendieta, 2013). However, the most widely used distinguishes two levels of organization: macrostructure and microstructure (Van Dijck & Kintsch, 1983; Vivas Vivas et al., 2021). Macrostructural analysis focuses on the structural components that constitute the organization of the episodes and the relationship between them (Heilmann et al., 2010; Stein & Glenn, 1979). In contrast, microstructural analysis examines language-dependent features, such as utterance length and syntactic complexity (Liles et al., 1995).

Finally, narrative production in retelling tasks can be assessed using two different approaches: norm-referenced and criterion-referenced (Crespo et al., 2015; Crespo, Figueroa-Leighton, et al., 2021; Vivas Vivas et al., 2021). The former evaluates retellings based on their adherence to a predefined model (Gillam & Pearson, 2004), while the latter focuses on differences in the arrangement, organization, and elaboration of narrative categories. Both approaches provide scores that distinguish performance protocols and enable monitoring of narrative skills (Strong, 1998).

Several assessment tools incorporate one or both of these approaches, including the Narrative Scoring Scheme (NSS) (Miller & Chapman, 2004); the Test of Narrative Language (Gillam & Pearson, 2004); the Profile of Oral Narrative Ability (PONA) (Westerveld & Gillon, 2010); the picture-based Multilingual Assessment Instrument for Narratives (LITMUS-MAIN) (Gagarina et al., 2012, 2015, 2019); and in Spanish, the *Evaluación del Desarrollo Narrativo* (EDNA) (Pavez, Coloma, et al., 2008), which includes a preliminary version of the Index of Narrative Complexity (INC) (Petersen et al., 2008). The INC evaluates the structural organization of oral narratives in retelling tasks by measuring the complexity of key narrative categories through a criterion-based approach (Petersen et al., 2008). In Chile, Bustos and Crespo (2014), followed by Crespo et al. (2015) and Crespo, Figueroa-Leighton, et al. (2021), refined the INC within the EDNA. Bustos & Crespo (2014) found that five-year-old children achieved success rates of over 50% in narrative production, particularly in character introduction, initiating events, attempts to resolve the problem, and consequences. However, they showed lower performance (below 30%) in internal responses and planning elements. Notably, Crespo et al. (2015) and Crespo, Figueroa-Leighton, et al. (2021) not only improved the methodological precision of the INC but also compared the narrative performance of typically developing (TD) children and children with Developmental Language Disorder (DLD). Their findings indicate that narrative complexity generally increased with age in both groups. Furthermore, there were differences between the TD and DLD groups that did not persist by the age of 10 (Crespo, Figueroa-Leighton, et al., 2021).

Narrative Performance in Children with Atypical Developmental Conditions: Developmental Language Disorder

The term Developmental Language Disorder (DLD) refers to a group of children who, despite not exhibiting severe sensory, neurological, emotional, intellectual, or behavioral impairments, exhibit deficits in particular linguistic and discourse abilities compared to their typically developing (TD) peers (Ervin, 2001). Developmental Language Disorder is considered a highly prevalent and heritable developmental condition (Bishop et al., 2017; Leonard, 2014), more frequently found in boys (Villanueva et al., 2011). However, the prevalence of DLD during childhood

development remains imprecisely defined. Studies conducted in Chile indicate that approximately 4% of children between the ages of three and seven are diagnosed with this disorder (De Barbieri et al., 1999). Additionally, reports show that by 2016, 18% of students enrolled in early childhood education (up to the age of 5 years and 11 months) were receiving special education support, with 95% of these cases corresponding to DLD (Chilean Ministry of Education [MINEDUC] & Centro de Estudios MINEDUC [CEM], 2018). In primary education (ages 6 to 13), the diagnosis of children with DLD has increased to nearly 75% between 2010 and 2016. Regarding gender differences, boys generally present a higher prevalence compared to girls, with some fluctuations over time. However, Granada-Azcárraga et al. (2020), who analyzed gender distribution in children aged 3 to 5 with DLD across three regions of Chile between 2004 and 2017, found that in two of these regions, enrollment rates for girls increased significantly, while they decreased among boys.

In terms of discourse and linguistic performance, children with DLD exhibit atypical development in both comprehension and oral or written expression compared to their TD peers (Leonard, 2014). In the case of Spanish-speaking children, significant difficulties have been reported in acquiring and mastering narrative discourse, particularly regarding the retrieval of the superstructure of texts, information recall, and the organization of causal structures during retelling tasks. These difficulties are accompanied by increased errors in both semantic and syntactic constructions (Acosta et al., 2012; Andreu et al., 2011; Auza et al., 2018; Coloma, 2014; Coloma et al., 2017; Crespo et al., 2015; del Valle Hernández et al., 2018; Pavez, Coloma, et al., 2008). Moreover, a large proportion of children aged 4 to 5 years display underdeveloped narrative skills (Coloma et al., 2002). Similarly, English-speaking children with DLD show difficulties in maintaining thematic coherence, sequencing events, and regulating the informativeness of episodes (Colozzo et al., 2011; Reuterskiöld et al., 2011; Squires et al., 2014). Regarding microstructural aspects, these children present impairments in linguistic productivity (Fey et al., 2004), utterance length (Vandewalle et al., 2012; M. F. Westerveld & Gillon, 2010), clause density (Colozzo et al., 2011; Fey et al., 2004), morphosyntactic measures (Colozzo et al., 2011; Reuterskiöld et al., 2011; Wetherell et al., 2007), and lexical diversity (Reuterskiöld et al., 2011; Squires et al., 2014).

Crespo and Figueroa (2016) reported differences when comparing the narrative performance of Spanish-speaking children with TD, DLD, and Intellectual Disability (ID). They analyzed the performance of a group of children with DLD matched with a group of TD children at age 5 using the Adapted Index of Narrative Complexity (INC-A) (Bustos & Crespo, 2014). Later, Crespo, Figueroa-Leighton, et al. (2021) examined the performance of TD and DLD children at ages 5 and 10. One of the study's objectives was to explore the developmental persistence or transient nature of DLD (Aguilar-Mediavilla et al., 2019; Conti-Ramsden et al., 2009; Ervin, 2001; Law et al., 2008; Sanz-Torrent et al., 2010). It is noteworthy that this is the only study to date employing a longitudinal methodology in this area. The results indicate that preschool children with DLD and TD predominantly favor causal chaining in their narratives, whereas children with ID tend to emphasize referential mention of characters.

Regarding the longitudinal follow-up, the researchers found that the inter-group differences observed at age 5 did not persist at age 10. Although improvements were reported in all indices, including the overall INC-A score for both groups across both time points, statistically significant differences were observed only in the group of children with DLD. These differences were found in all prototypical or essential criteria (Actions, Character, Initial Event, Plan, and Consequence) as well as in most of the non-prototypical criteria, except for Setting and Temporal Markers. The authors concluded that, although these results appear to support the hypothesis that describes DLD as a transient condition, such interpretation must be approached with caution, given the complexity of the disorder and the dimensions analyzed. Moreover, they speculate that "Temporal Markers," which function as structuring elements of narrative discourse, might be a key indicator for discriminating between children who have overcome the disorder and those who have not.

In summary, despite advances in refining instruments to assess children's narrative performance and the growing body of evidence—particularly regarding children with DLD—there is still no consensus as to which instrument or index is the most effective and sensitive for assessing narrative performance in this population. For instance, narrative skills assessments still fail to clearly differentiate between the cognitive, discourse, and linguistic skills required for narrative production and those required for retelling tasks (Kornev & Balčiūnienė, 2021; McCabe & Rollins, 1994). Moreover, it remains uncertain whether the INC can be reliably applied across different languages and cultural contexts (Balčiūnienė et al., 2019; Balčiūnienė & Kamandulytė-Merfeldienė, 2019; Vivas Vivas et al., 2021).

It is necessary to gather evidence regarding the sensitivity and psychometric properties of the INC. One of these properties is concurrent validity, referring to whether there is another measure

that can complement the information provided by the method in question. This study aims to address a methodological gap that has not been considered in previous research by complementing the analysis carried out by Crespo, Figueroa-Leighton, et al. (2021). To this end, we will compare the narrative performance of two groups of Chilean children, one with Developmental Language Disorder (DLD) and another with Typical Development (TD) matched by age, at two points in their development: at five and ten years of age. First, we will analyze whether the retellings produced by the participants are comparable in terms of quantitative or psycholinguistic measures in order to subsequently compare them based on qualitative criteria, specifically the INC-A. It is important to note that if significant inter-group differences are found in the quantitative measures, it would not be valid to conduct inter-group comparisons based on qualitative indices. Second, we will assess the reliability of the INC-A (Crespo, Figueroa-Leighton, et al., 2021) based on the independent evaluation of narrative performance by three coders.

Finally, we will explore whether the difficulties reported in the narrative performance of children with DLD are related to the development of structural narrative components or whether they can be explained by other measures (i.e., performance in other skills and/or functions encompassed within narrative competence, such as demands on episodic memory). To do so, we will compare inter-group performance using frequency statistics, and calculate differences in proportions to determine differences between the two groups at both time points.

The overall objective is to contribute to strengthening the INC so that it can be considered a valid tool for diagnosing and monitoring language disorders.

METHOD

Participants

A total of 39 children participated voluntarily in the study, after their parents signed an informed consent form, in accordance with the ethical principles outlined in the Declaration of Helsinki for research involving human participants (World Medical Association, 1975). The children were grouped according to their language development condition: those with Developmental Language Disorder (DLD) and those with Typical Development (TD) (see the characteristics of each group in Table 1). All participants were recruited from a longitudinal study that followed them from ages five through ten. For the present study, the performance of both groups was considered at two time points: at the beginning (age 5) (T1) and at the end (age 10) (T2). All the children attended schools in Viña del Mar and Valparaíso, some of which were semi-private (middle-class) and others publicly funded (lower-middle-class).

Table 1. Demographic characteristics of the DLD and TD participants.

	Gender		Mean at T1 (Age 5)	Mean at T2 (Age 10)
	Girls	Boys		
DLD	6	10	5.7	10.4
TD	11	12		

The participants were selected through convenience sampling (see details in Crespo et al., 2015; Crespo & Silva, 2019). In the first phase of the study (age five) several tests and indicators were administered. For the TD group, it was ensured that all children displayed age-appropriate cognitive and linguistic abilities and that their school attendance met institutional standards (i.e., the educational requirements of the schools they attended).

Children with DLD were selected based on institutional diagnoses carried out at the beginning of the academic year. The diagnosis was confirmed through a series of tests administered exclusively to this group: the Exploratory Test of Spanish Grammar by A. Toronto (STSG-R and STSG-E) (Pavez, 2003) and the Test to Assess Phonological Processes (TEPROSIF-R) (Pavez, Maggiolo, et al., 2008). In addition, these children underwent three specialized procedures: otoscopy, pure-tone audiometry, and speech audiometry, to rule out auditory impairments. The Raven's Coloured Progressive Matrices test (2005) was also administered to assess non-verbal cognitive abilities. Table 2 presents the average scores obtained on the STSG-E, STSG-R, TEPROSIF-R, and Raven tests, as well as the results of the pure-tone audiometry. These results show normal ranges in non-verbal cognitive abilities and hearing. Based on this information and clinical interviews, children were classified into two groups: TD and DLD.

Table 2. Results of the assessments for selection and diagnosis.

Tests			Pure-tone Audiometry		
	Mean	SD		Min	Max
STSG-E	23.4	6.64	Pure Tone Average AD	5	23.3
STSG-R	24.5	7.33			
TEPROSIF-R	30.9	14.2	Pure Tone Average AS	3.3	21.7
RAVEN	20.53	4.48			

Procedures and Materials

A total of 78 retellings were analyzed. The task that elicited these productions included three phases: first, each participant listened to an audiovisual story ("Flopi the Butterfly" at age 5 — T1 — and "The Sheep and the Extraterrestrial" at age 10 — T2). Then, they received a printed, picture-book-style version of the story containing sequenced images, which they could manipulate freely (Appendix 1). Finally, the participants were asked to retell the story to a second examiner who had not been involved in the previous phases. The children were allowed to use the picture book while retelling the story. The stories had a similar structure: both contained the same number of sequences in the same order, following the narrative grammar structure (Pavez, Coloma, et al., 2008; Stein & Glenn, 1979) (see Appendix 2).

Table 3. Total number of words, total number of clauses, and MLU (Mean Length of Utterance) in stimulus texts.

	Flopi The Butterfly (T1)	The Sheep and the Extraterrestrial (T2)
Number of Words	259	424
Number of Clauses	25	87
MLU	10.36	4.87

Data Analysis

All narratives were orthographically transcribed and segmented into clauses. A clause is defined as a segment of spoken discourse whose syntactic structure is composed of a verb/event and its dependent arguments (Crespo & Silva, 2019). The following measures were considered for the quantitative analysis (narrative length) of each retelling: total number of words, number of clauses, and Mean Length of Utterance (MLU). The qualitative analysis was based on the INC, adapted by Bustos & Crespo (2014) (see general description in Appendix 3), with modifications introduced by Crespo, Figueroa-Leighton, et al. (2021). Specifically, an evaluative component was incorporated into the narratives, and the scoring criteria for levels 2 and 3 were adjusted to reflect the same stage of development for the dimensions Dialogue Knowledge, Character, and Consequence (see INC-A details, Appendix 3). All narratives were assessed simultaneously and independently by three judges. A data analysis manual was developed, containing the INC-A along with guidelines for addressing problematic cases and providing examples.

Previous studies have identified notable performance differences across the various criteria that make up the INC (Balčiūnienė et al., 2019; Crespo et al., 2015; Crespo, Figueroa-Leighton, et al., 2021; Westerveld & Gillon, 2010). Therefore, we considered it necessary to isolate a set of core criteria (INC-core). In this study, the INC-core score was calculated as the mean of the sum of four criteria: Initial Event, Plan, Actions, and Consequences, along with one discourse performance criterion: Dialogue Introduction. The analyses included descriptive statistics (medians, ranges, and frequencies) and proportion difference tests.

To ensure the validity of the analysis, inter-rater agreement was assessed using the Kappa coefficient.

RESULTS

Regarding the instrument's efficacy, qualitative observations indicated that all children produced pragmatically appropriate retellings—that is, their responses met the task requirements and included the majority of the episodes from the original stories. These factors enhance the comparability of the elicited productions and, consequently, support the application of the INC-A. The analysis of inter-rater agreement, using the Kappa coefficient as a statistical measure of concordance proportion (Landis & Koch, 1977), yielded a Fleiss' Kappa coefficient of 0.601. This coefficient indicates a moderate level of agreement among the three evaluators (Landis & Koch, 1977).

A comparative analysis was conducted to address the comparability of performances between children with DLD and TD based on psycholinguistic measures.

Psycholinguistic Measures

Table 4 shows the analysis of the retellings' length and complexity.

Table 4. Extension: Mean number of words, clauses, and MLU in T1 and T2 by group.

	T1		T2	
Population	TD	DLD	TD	DLD
Number of Words	81	70	273	249
Number of Clauses	15	14	49	47
Utterance Length	1.1	0.9	0.5	0.6

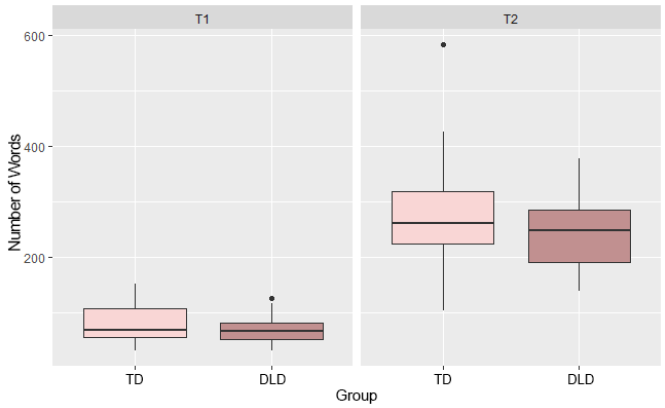

Figure 1. Extension of the narratives according to the number of words: Comparison between the TD and DLD groups.

Figure 1 shows that the median performance scores at the two assessment points are similar for both population groups. Additionally, a notable increase is observed at T2 in both the median values and the ranges. Despite this similarity, the TD group displays a broader range of performance at T1, with a median value slightly higher than that of the DLD group; conversely, the DLD group exhibits a noticeably more compact (homogeneous) performance. Although the TD group continues to show a broader range of performance with a slightly higher median at T2, the DLD group-level performance is more heterogeneous than that observed at T1.

In summary, the similarity of the medians confirms the comparability of productions between both groups across the two assessment points, with the TD group medians being slightly higher. The shift in performance pattern within the DLD group is noteworthy: while T1 exhibits homogeneous behavior, T2 displays a pattern more akin to that of the TD group.

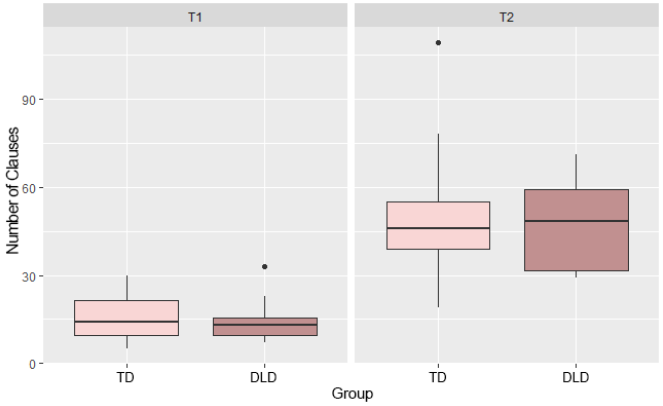

Figure 2. Extension of the narratives according to the number of clauses: Comparison between the TD and DLD group performances.

A similar performance pattern is observed in Figure 2 when considering the number of clauses. Although the medians of both groups (at T1 and T2) appear similar, and a notable increase from T1 to T2 is evident for both groups, some differences are recognized. While both groups display a compact range at T1, the degree of homogeneity in the DLD group is particularly notable. At T2, although both groups exhibit ranges with greater variance, the TD group shows a more homogeneous pattern compared to the DLD group. Additionally, the median values for the DLD group are slightly higher than those of the TD group.

Regarding the Mean Length of Utterance (MLU), shown in Figure 3, the previously described pattern changes. The median of the TD group is slightly higher than that of the DLD group at both time points. Furthermore, there is a marked increase in the median values between the two assessment points, with both groups displaying more homogeneous performance.

These analyses indicate no significant differences in the retellings of the two populations based on the selected psycholinguistic measures, confirming the feasibility of comparing INC analyses across both populations. On the other hand, the analyses reveal differences in performance patterns when considering T1 and T2, depending on the measure examined—namely, the number of words, the number of clauses, or the MLU. While the number of words and clauses follows a pattern from homogeneity to greater variance, the opposite is true for MLU.

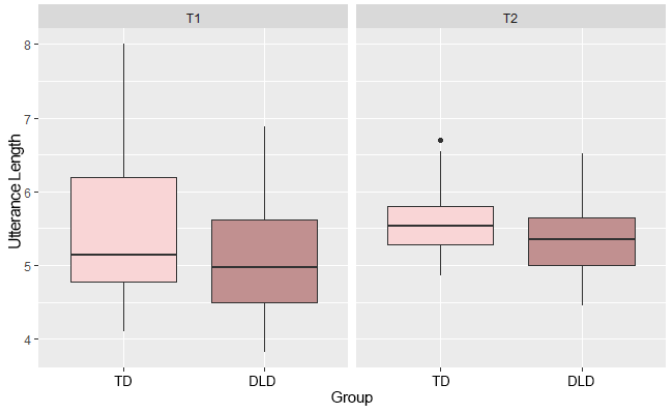

Figure 3. Extension of the narratives according to utterance length: Comparison between the TD and DLD group performances.

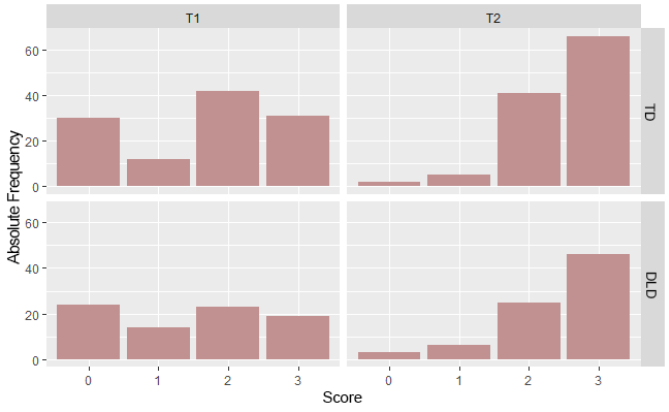

Figure 4. INC-A values: Frequency distribution chart of scores for TD and DLD groups at T1 and T2.

Index of Narrative Complexity (INC-A)

Below, we present the analysis of discourse performance based on the evaluation of the INC-A.

Table 5. INC scores and mean values in T1 and T2 for TD and DLD.

Time	T1		T2		INC
Population	Total	Mean	Total	Mean	
TD (N:23)	275	12	462	20	575
DLD (N:16)	175	11	312	20	400

The mean INC-A scores indicate that both groups achieved similar results at T1 and T2. Additionally, while both groups scored below 50% of the total score at T1 (43.7% for the DLD group and 47.8% for the TD group), both reached approximately 80% at T2.

Since the literature reports that children with DLD often omit episodes or present them incompletely, we decided to compare the population distribution according to INC-A scores (see Figure 4).

When comparing T1 and T2, both groups exhibit similar developmental trends, characterized by a reduction in absent mentions (score 0) and inappropriate mentions (scores 1 and 2), alongside an increase in adequate mentions (score 3). However, it is necessary to determine whether these differences are statistically significant. A hypothesis test for proportion differences was conducted with a significance level (α) of 0.05. The analysis revealed that, for the TD group, statistically significant differences occurred only at scores 0 ($p = 0.000269$) and 3 ($p = 0.00563$), while no significant differences were found for scores 1 ($p = 0.1305$) and 2 ($p = 1$). Similarly, significant differences were observed in the DLD group only for scores 0 ($p = 0.0243$) and 3 ($p = 0.0285$), with no significant differences for scores 1 ($p = 0.094$) and 2 ($p = 0.863$).

In summary, both populations display similar developmental trends: significant increases in adequate mentions correspond with significant decreases in instances where children omitted the evaluated category. Given that one of the study's objectives is to identify whether differences exist between groups at each assessment point, we further analyzed significant differences across the different scores. We conducted a proportion difference test with a significance level of $\alpha = 0.05$, and the results are presented in Table 6.

Table 6. Differences in proportions between TD and DLD groups based on INC-A scores at ages 5 and 10.

Scoring Scale	T1 (Age 5)	T2 (Age 10)
0	0.6614	0.6794
1	0.225	0.5333
2	0.3281	0.5424
3	0.7356	0.999

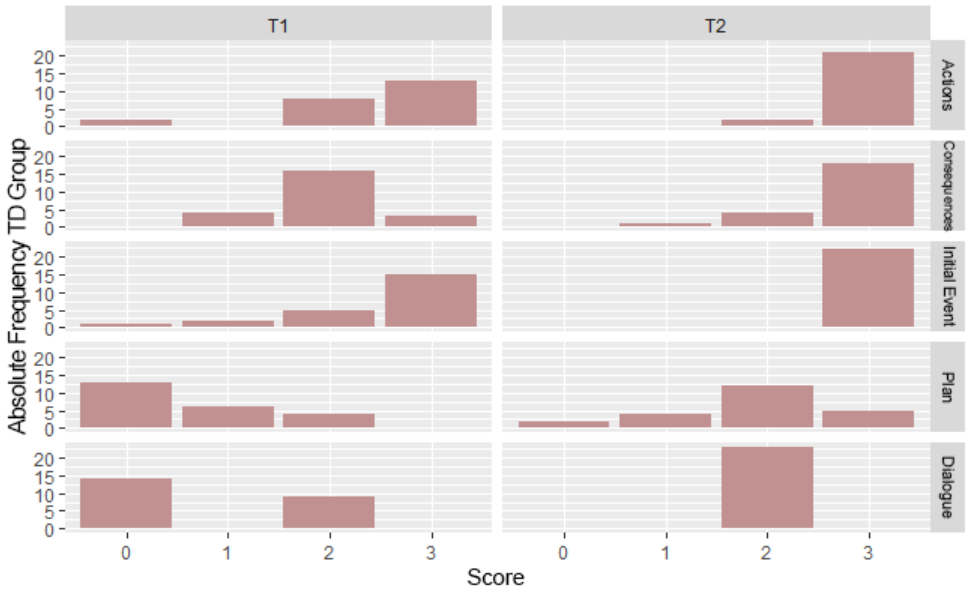

Figure 5. Comparison of performance on core criteria at T1 and T2 based on INC-A scores: TD group

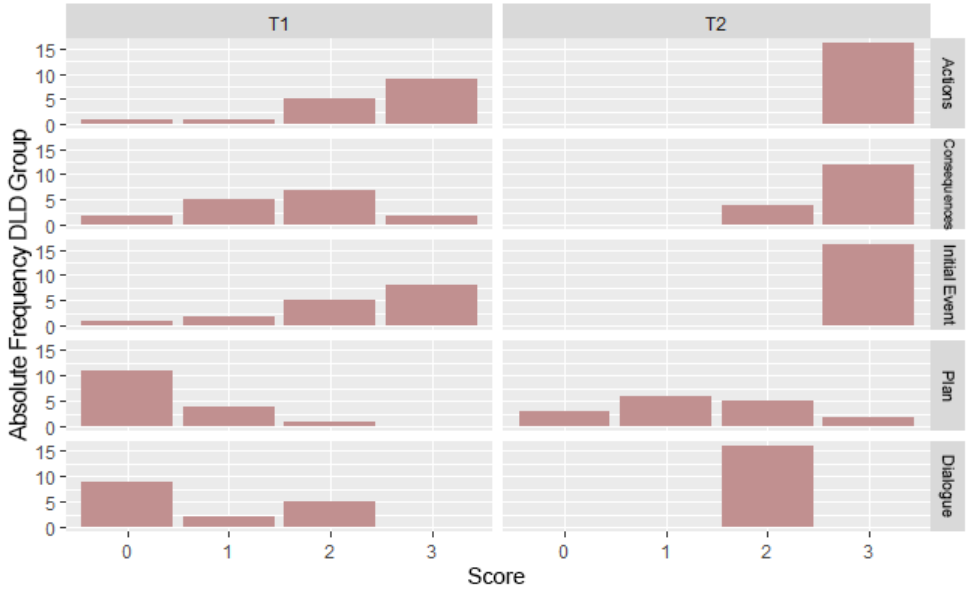

Figure 6. Comparison of performance on core criteria at T1 and T2 based on INC-A scores: DLD group.

The analysis yielded no significant differences between the DLD and TD groups in any of the core INC scores at any assessment point. However, despite this finding, we analyzed the differences across the various criteria.

The analysis of specific criteria, presented in Figures 6 and 7, revealed a similar developmental trend for both the TD and DLD groups in several core INC components (Actions, Consequences, Initial Event, and Dialogue Introduction): at T1 performance tends to be classified as absent, incomplete, or erroneous (scores 0, 1, and 2), whereas at T2, the scores concentrate at 3. The exception is the *Plan* criterion, where this trend is observed only in the TD group; in contrast, the DLD group continues to show absent or incomplete mentions of the episode even at age 10. These findings suggest that the *Plan* criterion may serve as a distinguishing feature between TD and DLD performance, although confirming this assumption requires a formal difference analysis.

To qualitatively represent the reported difference, we show four excerpts from the *Plan* sequence uttered by both populations, allowing us to examine the linguistic strategies each group of children employs to construct this episode.

Table 7. Analysis of differences and similarities in the Plan sequence structure according to population and elicitation context (T1 and T2).

Population	Original Text	TD	DLD
T1	*He trapped her inside a jar. Flopi was very sad and scared because she could neither fly nor move.* ***Her worried friends started talking among themselves, asking, "What should we do? We must save her no matter what! Let's wait until Mr*** **Bigotes** ***falls asleep, and then we'll rescue her."***	*He put her inside a jar and sealed it. Flopi couldn't fly and was very sad. Then, he closed the door, that could not be opened.* ***With her friends, they said, "We have to save Flopi; we must wait until the grumpy man falls asleep."*** *Until he falls asleep, they opened the lid. They promised never to go there again.* (Flopi, 52 TD)	*The butterfly Flopi and they were happy and went to a field and first the man caught her caught her and took her to his house and she couldn't move.* ***First, the two said, "When he falls asleep, we trap her to get out,"*** *and they did so happily and it ended.* (Flopi, 18 DLD)
T2	*Otilia looked at Frido and felt very sorry for him. She thought that, besides the cold, he must feel sad and scared.* ***So, she decided to help him. "I will lend him some of my wool to keep warm," the sheep thought. "We could be friends."***	*A Martian appeared and said that he was very cold but he doesn't need wool, so the sheep felt sad for him and* ***thought about making him a sweater.*** *Then with her skills she began knitting him a coat.* (Otilia, 57 TD)	*The extraterrestrial was cold and needed something to cover himself and* ***the sheep thought "Couldn't I give him my wool?"*** *and then she imagined giving him her wool and wrapping him.* (Otilia, 27 DLD)

A comparison of the elicited Plan sequences reveals that at T1, children in the TD group, although sometimes lexically imprecise (e.g., referring to the "door" instead of the "jar lid" in which the butterfly is trapped), appropriately present the conflict, referentially introduce the characters involved in the Plan, and explicitly state the Plan itself, using the precise opposite of the antagonist's action: *"We have to save Flopi."* They even identify the difficulty of carrying out the Plan (*"the grumpy man"*). In contrast, children with DLD, although they identify the conflict (*"Flopi has been trapped"*), do not mention the container (*"the jar"*), which makes it challenging to understand the event in the final utterance: *"We are going to trap her so she can escape."* The butterflies are supposed to free Flopi, but the continued lexical use of the term *"trap"* interferes with the proper construction of the sequence. At T2, although the elicited sequence reveals that the child recalls the conflict and introduces a strategy to resolve it, the narrative is more concise, resembling a report of events rather than a narrative from a narrator's perspective (Fina, 2021).

DISCUSSION

Instrument: Properties

Regarding the instrument, providing a story for reference reduces the attentional and planning demands on the child, unlike those

that rely on picture description tasks to elicit narratives (e.g., the NSS by Miller & Chapman, 2004). In this task, the child retells the story to a communication partner who was not present during the original narration, creating a communicative situation that reflects real-world pragmatic demands (ecological validity). Narrating a story for someone who did not hear it is inherently more natural than producing a narrative for an evaluative task (Wofford et al., 2022). This approach minimizes the risk of narrative omissions or "narrative erosion" and further reinforces the ecological validity of the task (Channell et al., 2018). Moreover, during the retelling, children can manipulate images, which helps reduce the cognitive load on episodic and working memory, freeing up cognitive resources for discourse planning and formulation. This design allows children with DLD to produce narratives more fluently, as the linguistic demands do not compete with other cognitive processes, such as memory. This factor may partly explain the differences observed between our results and those reported by Maggiolo et al. (2003) and Fey et al. (2004).

In this vein, Wofford et al. (2022) compared microstructural measures (lexical diversity, mean length of utterance, and percentage of grammatical utterances) within the same child population using two elicitation techniques: picture-based storytelling and retelling of a heard story. Their findings revealed no significant differences between the two tasks. However, in line with previous studies (Lucero & Uchikoshi, 2019; Schneider & Dubé, 2005), they observed that retellings tend to be structurally more complete and lexically richer (Westerveld & Gillon, 2010). We believe this factor also contributed to our results. Finally, when comparing our findings with those of Pavez, Coloma, et al. (2008) as well as Bustos & Crespo (2014), Crespo et al. (2015), and Crespo, Figueroa-Leighton, et al. (2021), we consider that modifications to the design and administration conditions influenced the quality of the data. Additional contributing factors include adapting the stimulus texts to the children's age group and designing the image sequences. The narratives incorporate conflicts in which characters' motivations are conveyed through emotions and thoughts, making the stories more engaging for children.

Regarding the internal validity of the INC-A, although the inter-rater agreement index reached a moderate level, it is considered acceptable given the number of weighted criteria and the high intrinsic variability of the rating scales. We observed that discrepancies were greater for some criteria than for others, providing guidance on which aspects should be further clarified during the scoring process.

In summary, the instrument is efficient and methodologically rigorous, as reflected in its high internal and external validity. However, its level of concurrent validity remains to be assessed (Castejón Costa, 1997). Furthermore, the coherence and systematic modifications introduced by Crespo et al. (2015) and Crespo, Figueroa-Leighton, et al. (2021) have substantially improved the instrument, consolidating its psychometric properties and enhancing its predictive validity and clinical potential. In line with previous studies (Crespo et al., 2015; Crespo, Figueroa-Leighton, et al., 2021), we consider it necessary to complement the instrument with additional distinctions related to microstructural criteria (e.g., levels of explicitness in the Plan sequence, gradations in the use of temporal markers, etc.), which could also be proposed as developmental indicators.

Narrative Performance: DLD and TD Groups

Regarding performance, the productivity measures at two developmental time points revealed that, in both groups, not only do score values increase over time, but there are also modifications in group performance patterns. In this regard, Crespo, Silva, et al. (2021) observed that the dynamics of development are shaped by intragroup differences that gradually increase and are responsible for driving overall group performance. Our data further support this notion, highlighting the need to study the specific characteristics of these differences. Such knowledge would enable, for example, the design of targeted clinical interventions. By identifying specific characteristics, we can develop tools that promote these aspects to foster the development of children with lower performance levels.

In line with the objectives of our study, we observed a notable increase in total and core INC-A scores at both assessment points. This increase was approximately 40% in both groups. A disaggregated analysis of the population frequency distributions revealed that, at the age of five, both groups produced narratives with absent category introductions. In contrast, by age ten, both groups produced narratives that included all evaluated categories with a high level of adequacy. They also showed a similar performance pattern: the only scoring categories in which significant differences were observed between the two assessment points were those qualifying absent mentions as well as complete and appropriate mentions. These findings complement the developmental characterization presented in Crespo et al. (2015) and Crespo, Figueroa-Leighton, et al. (2021).

Regarding performance differences between the DLD and TD groups, our analysis—unlike previous studies (Crespo et al., 2015; Crespo, Figueroa-Leighton, et al., 2021; Leonard, 2014;

Maggiolo et al., 2003) —found no significant differences at either assessment point across any of the scores. The only difference was found in the evaluation of performance at age ten, specifically regarding the Plan criterion (a structural criterion).

Does this imply that the narrative production of children with DLD is not impaired, as suggested by previous research (Bustos & Crespo, 2014; Crespo et al., 2015; Maggiolo et al., 2003; Pavez, Coloma, et al., 2008)? Our results differ from those of other studies (Maggiolo et al., 2003; Pavez, Coloma, et al., 2008) due to methodological modifications in our research design. These include adaptations made to the INC (Bustos & Crespo, 2014; Crespo et al., 2015; Crespo, Figueroa-Leighton, et al., 2021), as well as the decision to report data from a longitudinal study, which inherently reduces intragroup variability. At the same time, our findings are consistent with those of Crespo et al. (2015) and Crespo, Figueroa-Leighton, et al. (2021) in that neither the TD nor the DLD group reached the maximum score at age five. Our data confirm their speculative projection regarding the development of narrative categories, as by age ten, children in both groups achieved 80% of the maximum score. The main differences with Crespo, Figueroa-Leighton, et al. (2021) stem from decisions regarding data analysis and processing. The need to enhance the psychometric validity of the instrument led to the use of independent coders for score assignment, enabling the obtention of interrater reliability values.

Furthermore, analyzing proportional differences rather than means made it possible to better capture each group's performance relative to the specific distribution of frequency scores. In this regard, differences were observed in identifying which category best discriminates between performances. While Crespo, Figueroa-Leighton, et al. (2021) focused on the Character category, our analyses suggest that the Plan category is more sensitive to group differences. This category is possibly more challenging for children with DLD, as it requires them to apply mentalistic abilities (Baron-Cohen, 2000). Qualitative analyses revealed that, although shortcomings identified at age five may be partially resolved, by age ten, the DLD group continues to produce narratives with limited narrative positioning, resembling more a report of events (Fina, 2021). These differences in performance across the two populations may reflect varying degrees of mastery over the discourse, grammar, and cognitive frameworks that underpin skilled narrative construction (Fina, 2021).

CONCLUSIONS

This study provides evidence regarding the psychometric properties of an instrument to evaluate the performance and development of narrative skills in Spanish-speaking children. Although some of the results contradict findings from previous research, this study contributes to the growing body of evidence on the developmental trajectory of narrative skills in both TD and DLD children.

The instrument provided a corpus with a high degree of ecological validity, allowing for the analysis of narrative performance through both quantitative linguistic productivity measures and specific qualitative indicators.

Moreover, the results suggest new areas for inquiry, such as the relationship between the development of narrative skills and cognitive abilities related to information storage, organization, and retrieval. In this respect, it is essential to complement narrative assessment with specific measures designed to evaluate these cognitive domains. Finally, research such as this one represents a significant contribution to the development of assessment and intervention tools for clinical settings.

REFERENCES

Acosta, V., Moreno, A., & Axpe, Á. (2012). Intervención logopédica sobre habilidades narrativas en niños con Trastorno Específico del Lenguaje. *Journal for the Study of Education and Development*, *35*(2), 201–213. https://doi.org/10.1174/021037012800218032

Aguilar-Mediavilla, E., Buil-Legaz, L., López-Penadés, R., Sanchez-Azanza, V. A., & Adrover-Roig, D. (2019). Academic Outcomes in Bilingual Children With Developmental Language Disorder: A Longitudinal Study. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00531

Andreu, L., Sanz-Torrent, M., Guàrdia Olmos, J., & Macwhinney, B. (2011). Narrative comprehension and production in children with SLI: An eye movement study. *Clinical Linguistics & Phonetics*, *25*(9), 767–783. https://doi.org/10.3109/02699206.2011.565542

Asociación Médica Mundial. (1975, octubre). *Declaración de Helsinki. Principios éticos para las investigaciones con seres humanos*. Normas y Documentos de Etica Médica. Colegio Médico de Chile (AG). http://23118.psi.uba.ar/academica/carrerasdegrado/psicologia/informacion_adicional/practicas_de_investigacion/775/otros_documentos/DECLARACION%20DE%20HELSINKI.pdf

Auza, A., Harmon, M. T., & Murata, C. (2018). Retelling stories: Grammatical and lexical measures for identifying monolingual Spanish speaking children with specific language impairment (SLI). *Journal of Communication Disorders*, *71*, 52–60. https://doi.org/10.1016/j.jcomdis.2017.12.001

Balčiūnienė, I., & Kamandulytė-Merfeldienė, L. (2019). Narrative structure and language in specifically language-impaired children. *Typical and Atypical*

Language Development Symposium. https://www.hse.ru/mirror/pubs/share/303591024

Balčiūnienė, I., Kornev, A. N., & Aguilar-Mediavilla, E. (2019). Evaluation of narrative skills in language-impaired children: Advantages of a dynamic approach. En L. Buil-Legaz, R. López-Penadés, V. Sánchez-Azanza, & D. Adrover-Roig (Eds.), *Atypical Language Development in Romance Languages* (pp. 127–141). John Benjamins Publishing Company. https://benjamins.com/catalog/z.223

Baron-Cohen, S. (2000). Theory of mind and autism: A review. *International Review of Research in Mental Retardation*, *23*, 169–184. https://doi.org/10.1016/S0074-7750(00)80010-5

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

Bruner, J. (1986). *Realidad mental y mundos posibles: Los actos de la imaginación que dan sentido a la experiencia* (B. López, Trad.). Gedisa Editorial.

Bustos, A., & Crespo, N. M. (2014). Comprensión oral de narraciones y producción narrativa: Dos medidas a través de una tarea de recontado. *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile*, *30*, 7. https://dialnet.unirioja.es/servlet/articulo?codigo=6354604

Castejón Costa, J. L. (1997). *Introducción a los métodos y técnicas de investigación y obtención de datos en psicología*. Editorial Club Universitario (ECU). https://dialnet.unirioja.es/servlet/libro?codigo=105294

Channell, M. M., Loveall, S. J., Conners, F. A., Harvey, D. J., & Abbeduto, L. (2018). Narrative Language Sampling in Typical Development: Implications for Clinical Trials. *American Journal of Speech-Language Pathology*, *27*(1), 123–135. https://doi.org/10.1044/2017_AJSLP-17-0046

Coloma, C. J. (2014). Discurso narrativo en escolares de 1° básico con Trastorno Específico del Lenguaje (TEL). *Revista signos*, *47*(84), 3–20. https://doi.org/10.4067/S0718-09342014000100001

Coloma, C. J., Mendoza, E., & Carballo, G. (2017). Desempeño gramatical y narrativo en niños con Trastorno Específico del Lenguaje. *ResearchGate*, *69*, 67–90. https://doi.org/10.5209/CLAC.55314

Coloma, C. J., Pavez, M. M., & Maggiolo L., M. (2002). Caracterización, análisis y estimulación del discurso narrativo en niños con trastorno específico del lenguaje. *Revista Chilena de Fonoaudiología*, *3*, 75–90. https://pesquisa.bvsalud.org/portal/resource/pt/lil-313307

Colozzo, P., Gillam, R. B., Wood, M., Schnell, R. D., & Johnston, J. R. (2011). Content and Form in the Narratives of Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, *54*(6), 1609–1627. https://doi.org/10.1044/1092-4388(2011/10-0247)

Conti-Ramsden, G., Durkin, K., Simkin, Z., & Knox, E. (2009). Specific language impairment and school outcomes. I: Identifying and explaining variability at the end of compulsory education. *International Journal of Language & Communication Disorders*, *44*(1), 15–35. https://doi.org/10.1080/13682820801921601

Crespo, N., & Figueroa, A. (2016). The different narrative profiles of children with different linguistic and cognitive conditions. *Literatura y lingüística*, *33*, 443–464. https://doi.org/10.4067/S0716-58112016000100021

Crespo, N., Figueroa-Leighton, A., & Góngora, B. (2021). Índice de Complejidad Narrativa Adaptado en escolares chilenos con y sin historia de trastorno específico del lenguaje. *Logos: Revista de Lingüística, Filosofía y Literatura*, *31*(2), Article 2. https://doi.org/10.15443/RL3120

Crespo, N., Góngora, B., & Figueroa-Leighton, A. (2015). La medición por criterio de la narrativa en niños con trastorno específico del lenguaje. *Revista de Logopedia, Foniatría y Audiología*, *35*(3), 114–122. https://doi.org/10.1016/j.rlfa.2014.12.003

Crespo, N., & Silva, M. L. (2019). Cuando las palabras del otro se hacen mías: Discurso reportado y flexibilidad pragmática. En niños con desarrollo típico y con trastorno específico de lenguaje. *Lingüística*, *35*(2), 215–234. https://doi.org/10.5935/2079-312x.20190025

Crespo, N., Silva, M. L., Sepúlveda Toro, J. I., Rubbo, Y., Garay Frontini, M. B., & Tenaglia, M. F. (2021). Reporte de dichos y pensamientos en niños con trastorno del desarrollo del lenguaje: Rasgos evolutivos del discurso indirecto. *Signos Lingüísticos*, *17*(33), Article 33. https://signoslinguisticos.izt.uam.mx/index.php/SL/article/view/307

De Barbieri, Z., Maggiolo L., M., & Alfaro Y., S. (1999). Trastornos de la comunicación oral en niños que asisten a control de salud en un consultorio de atención primaria. *Revista Chilena de Pediatría, 70*(1), 36–40. https://pesquisa.bvsalud.org/portal/resource/pt/lil-244008

del Valle Hernández, G., Acosta Rodríguez, V. M., Ramírez Santana, G. M., del Valle Hernández, G., Acosta Rodríguez, V. M., & Ramírez Santana, G. M. (2018). The grammatical production in the narrative discourse of pupils with Specific Language Impairment (SLI). *Revista Signos*, *51*(98), 264–284. https://doi.org/10.4067/S0718-09342018000300264

Ervin, M. (2001). SLI: What We Know and Why It Matters. *The ASHA Leader Archive*, *6*(12), 4–31. https://doi.org/10.1044/leader.FTR1.06122001.4

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and Written Story Composition Skills of Children With Language Impairment. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1301–1318. https://doi.org/10.1044/1092-4388(2004/098)

Fina, A. D. (2021). Doing narrative analysis from a narratives-as-practices perspective. *Narrative Inquiry*, *31*(1), 49–71. https://doi.org/10.1075/ni.20067.def

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. En S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (pp. 243–276). Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781783093137

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2012). ZAS Berlin: MAIN Start. *Zas Papers in Linguistics*, *56*, 1–140. https://www.leibniz-zas.de/de/publications/schriftenreihe/zaspil/zaspil-56/main-start/

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics*, *63*, 1–21. https://doi.org/10.21248/zaspil.63.2019.516

Gillam, R. B., & Pearson, N. A. (2004). *Test of narrative language*. Pro-Ed.

Granada-Azcárraga, M., Cáceres Zúñiga, M. F., Pomés Correa, M. P., & Ibáñez Córdova, A. (2020). Tendencia de matrícula de estudiantes con TEL en escuelas especiales de lenguaje en Chile. *RLA. Revista de Lingüística Teórica y Aplicada*, *58*(1), Article 1. https://doi.org/10.29393/RLA58-5MGTM40005

Griffith, P. L., Ripich, D. N., & Dastoli, S. L. (1986). Story structure, cohesion, and propositions in story recalls by learning-disabled and nondisabled children. *Journal of Psycholinguistic Research*, *15*(6), 539–555. https://doi.org/10.1007/BF01067635

Gummersall, D. M., & Strong, C. J. (1999). Assessment of Complex Sentence Production in a Narrative Context. *Language, Speech, and Hearing Services in Schools*, *30*(2), 152–164. https://doi.org/10.1044/0161-1461.3002.152

Heilmann, J., Miller, J. F., & Nockerts, A. (2010). Sensitivity of narrative organization measures using narrative retells produced by young school-age children. *Language Testing*, *27*(4), 603–626. https://doi.org/10.1177/0265532209355669

Kornev, A. N., & Balčiūnienė, I. (2021). Lexical and Grammatical Errors in Developmentally Language Disordered and Typically Developed Children: The Impact of Age and Discourse Genre. *Children*, *8*(12), Article 12. https://doi.org/10.3390/children8121114

Landis, J. R., & Koch, G. G. (1977). An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics*, *33*(2), 363–374. https://doi.org/10.2307/2529786

Law, J., Tomblin, J. B., & Zhang, X. (2008). Characterizing the Growth Trajectories of Language-Impaired Children Between 7 and 11 Years of Age. *Journal of Speech, Language, and Hearing Research*, *51*(3), 739–749. https://doi.org/10.1044/1092-4388(2008/052)

Leonard, L. B. (2014). *Children with Specific Language Impairment* (2ª ed.). The MIT Press. https://doi.org/10.7551/mitpress/9152.001.0001

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of Narrative Discourse Ability in Children With Language Disorders. *Journal of Speech, Language, and Hearing Research*, *38*(2), 415–425. https://doi.org/10.1044/jshr.3802.415

Lucero, A., & Uchikoshi, Y. (2019). Narrative Assessments with First Grade Spanish-English Emergent Bilinguals: Spontaneous versus Retell Conditions. *Narrative Inquiry: NI*, *29*(1), 137–156. https://doi.org/10.1075/ni.18015.luc

Maggiolo, M., Pavez, M. M., & Coloma, C. J. (2003). Terapia para el desarrollo narrativo en niños con trastorno específico del lenguaje. *Revista de Logopedia, Foniatría y Audiología*, *23*(2), 98–108. https://doi.org/10.1016/S0214-4603(03)75750-7

Masterson, J. J., & Kamhi, A. G. (1991). The Effects of Sampling Conditions on Sentence Production in Normal, Reading-Disabled, and Language-Learning-Disabled Children. *Journal of Speech, Language, and Hearing Research*, *34*(3), 549–558. https://doi.org/10.1044/jshr.3403.549

McCabe, A., & Rollins, P. R. (1994). Assessment of Preschool Narrative Skills. *American Journal of Speech-Language Pathology*, *3*(1), 45–56. https://doi.org/10.1044/1058-0360.0301.45

Mendieta, J. A. (2013). Narrative research: An alternative approach to study language teaching and learning. *Folios*, *37*, 135–147. https://doi.org/10.17227/01234870.37folios135.147

Miller, J. F., & Chapman, R. S. (2004). *Systematic analysis of language transcripts* (Versión 8) [Software]. Language Analysis Laboratory, Waisman Center, University of Wisconsin-Madison. https://www.saltsoftware.com/

Ministerio de Educación [MINEDUC], & Centro de Estudios MINEDUC [CEM]. (2018). *Indicadores de la educación en Chile 2010-2016.* http://bibliotecadigital.mineduc.cl//handle/20.500.12365/2405

Pavez, M. M. (2003). *Test exploratorio de gramática española de A. Toronto. Aplicación en Chile*. Ediciones UC. http://ediciones.uc.cl/ediciones-uc-educaci%C3%B3n-test-exploratorio-gramatica-espanola-toronto-p-468.html

Pavez, M. M., Coloma, C. J., & Maggiolo, M. (2008). *El desarrollo narrativo en niños. Una propuesta práctica para la evaluación y la intervención en niños con trastorno del lenguaje* (No. 19). ARS Médica. https://rchd.uc.cl/index.php/onom/article/view/32077

Pavez, M. M., Maggiolo, M., & Coloma, C. J. (2008). *Test para evaluar procesos de simplificación fonológica: TEPROSIF-R*. Ediciones UC.

Petersen, D. B., Gillam, S. L., & Gillam, R. B. (2008). Emerging Procedures in Narrative Assessment: The Index of Narrative Complexity. *Topics in Language Disorders*, *28*(2), 115. https://doi.org/10.1097/01.TLD.0000318933.46925.86

Raven, J. C. (2005). *Test de matrices progresivas: Escala coloreada : cuaderno de matrices : series A, Ab y B* (1a. ed.). Paidós.

Reuterskiöld, C., Hansson, K., & Sahlén, B. (2011). Narrative skills in Swedish children with language impairment. *Journal of Communication Disorders*, *44*(6), 733–744. https://doi.org/10.1016/j.jcomdis.2011.04.010

Romero Contreras, S., & Gómez Martínez, G. E. (2013). El desarrollo del lenguaje evaluativo en narraciones de niños mexicanos de 3 a 12 años. *Actualidades en Psicología*, *27*(115), 15–30. https://doi.org/10.15517/ap.v27i115.8674

Sanz-Torrent, M., Andreu, Ll., Badia, I., & Serra, M. (2010). El proceso lector en niños con antecedentes de retraso y trastorno específico del lenguaje. *Revista de Logopedia, Foniatría y Audiología*, *30*(1), 23–33. https://doi.org/10.1016/S0214-4603(10)70005-X

Schneider, P., & Dubé, R. V. (2005). Story presentation effects on children's retell content. *American Journal of Speech-Language Pathology*, *14*(1), 52–60. https://doi.org/10.1044/1058-0360(2005/007)

Silva, M. L. (2008). *Desarrollo y gramaticalización de las claúsulas relativas en el discurso infantil: Aportes desde el enfoque cognitivo-prototípico* [Tesis Doctoral, Universidad Nacional de La Plata]. https://163.10.30.7/Record/TESIS00782

Silva, M. L. (2010). El estudio de la sintaxis infantil a partir del diálogo con niños: Aportes metodológicos. *Interdisciplinaria*, *27*(2), 277–296. https://www.scielo.org.ar/scielo.php?script=sci_abstract&pid=S1668-70272010000200006&lng=es&nrm=iso&tlng=es

Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language & Communication Disorders*, *49*(1), 60–74. https://doi.org/10.1111/1460-6984.12044

Stein, N., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. En R. O. Freedle (Ed.), *Advances in discourse, new directions in discourse processing processes* (pp. 53–120). Ablex Publishing Corporation.

Strong, C. J. (1998). *The Strong Narrative Assessment Procedure*. Thinking Publications.

Vandewalle, E., Boets, B., Boons, T., Ghesquière, P., & Zink, I. (2012). Oral language and narrative skills in children with specific language impairment with and without literacy delay: A three-year longitudinal study. *Research in Developmental Disabilities*, *33*(6), 1857–1870. https://doi.org/10.1016/j.ridd.2012.05.004

Villanueva, P., Newbury, D. F., Jara, L., De Barbieri, Z., Mirza, G., Palomino, H. M., Fernández, M. A., Cazier, J.-B., Monaco, A. P., & Palomino, H. (2011). Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. *European Journal of Human Genetics*, *19*(6), 687–695. https://doi.org/10.1038/ejhg.2010.251

Vivas Vivas, S., Garay Frontini, M., & Silva, M. L. (2021). Renarración infantil de cuentos ficcionales: Incidencia de la temática en las estructuras narrativas. En K. H. Zimmermann & L. J. A. Neve (Eds.), *Desarrollo lingüístico tardío en poblaciones hispanohablantes* (pp. 147–179). Comunicacion Científica. https://www.google.cl/books/edition/Desarrollo_ling%C3%BC%C3%ADstico_tard%C3%ADo_en_pobl/xbLdEAAAQBAJ?hl=es&gbpv=0

Volkmar, F. R. (2020). *Encyclopedia of Autism Spectrum Disorders*. Springer New York.

Westerveld, M. F., & Gillon, G. T. (2010). Profiling oral narrative ability in young school-aged children. *International Journal of Speech-Language Pathology*, *12*(3), 178–189. https://doi.org/10.3109/17549500903194125

Westerveld, M., & Gillon, G. (1999). Narrative language sampling in young school-age children. *New Zealand Journal of Speech and Language Therapy*, *53/54*, 34–41. https://www.researchgate.net/publication/305770376_Narrative_language_sampling_in_young_school-age_children

Wetherell, D., Botting, N., & Conti-Ramsden, G. (2007). Narrative in adolescent specific language impairment (SLI): A comparison with peers across two different narrative genres. *International Journal of Language & Communication Disorders*, *42*(5), 583–605. https://doi.org/10.1080/13682820601056228

Wofford, M. C., Cano, J., Goodrich, J. M., & Fitton, L. (2022). Tell or Retell? The Role of Task and Language in Spanish–English Narrative Microstructure Performance. *Language, Speech, and Hearing Services in Schools*, *53*(2), 511–531. https://doi.org/10.1044/2021_LSHSS-21-00055

APPENDIX 1. Sample sequence of images from stimulus stories

Story "Flopi the Butterfly"

Story "The Sheep and the Extraterrestrial"

APPENDIX 2. Stimulus texts with sequences of the story grammar.

	Story Grammar of the Tale "Flopi the Butterfly"
Scene Establishment	Once upon a time, there was a butterfly named Flopi. Flopi loved to dance in the wind and go out with her friends. Every day, they would stroll through the park and have lots of fun singing and laughing.
Initial Event	One day, Flopi and her friends happily went flying through the field full of flowers and fruit. The field belonged to a very grumpy man named Mr. Bigotes. They were flying over the daisies when, suddenly, they felt that someone was watching them from behind a tree.
Internal Response	It was Mr. Bigotes, annoyed that the butterflies were playing in his field!
Plan	I'm going to hunt them, he thought.
Attempt	He went to his house and looked for his butterfly net. Once back in the field, he began to chase the butterflies. Then Flopi and her friends started flying faster and faster.
Direct Consequence	But Flopi wasn't able to escape and was caught by Mr. Bigotes.
Initial Event	He took her to his home and trapped her inside a jar.
Internal Response	Flopi was very sad and scared because she could neither fly nor move.
Plan	Her worried friends started talking among themselves, asking, "What should we do? We must save her no matter what! Let's wait until Mr. Bigotes falls asleep, and then we'll rescue her."
Attempt	Mr. Bigotes went to sleep, and the butterfly friends flew in through the window into the house. Together, they twisted the jar lid and freed Flopi.
Direct Consequence	Flopi and her friends quickly fled Mr. Bigotes' house and flew away to get far from the field.
Reaction and Resolution	The butterflies were very happy and promised never to return to that place again.

	Story Grammar of the Tale "The Sheep and the Extraterrestrial"
Scene Establishment	Once upon a time, there was a sheep named Otilia who lived in a meadow. The other sheep were her friends, and they all played and chatted while grazing on the tender grass. It was winter and very cold, but she didn't worry at all. Her body was covered with white wool, and she was very warm.
Initial Event	One night, Otilia was grazing alone in the meadow while the other sheep slept under a tree. Suddenly, she saw a very bright light in the sky—it was a giant spaceship flying overhead. Suddenly, the little sheep realized she was floating in the air until she entered through one of the spaceship's doors.
Internal Response	Otilia was terrified; she had no idea what was going to happen to her. She began to cry and scream, "Help! I need someone to help me!"
Plan	"I'll try to escape," she thought. "I don't want to stay on this ship."
Attempt	Otilia pushed the door of the spaceship to escape. But her wool got caught on the handle. "Ohh," the sheep lamented, "I won't be able to get out."
Initial Event	Suddenly, a green extraterrestrial appeared, with big eyes and a pointed mouth. "My name is Frido," he exclaimed. "Don't be afraid, I won't hurt you. I only want you to share some of your wool with me, because I'm cold on this planet, and besides, I feel lonely because I have no friends."
Internal Response	Otilia looked at Frido and felt very sorry for him. She thought that, besides being cold, he must also feel sad and scared.

Plan	So, she decided to help him. “I will lend him some of my wool so he can keep warm,” thought the sheep. “We could be friends.”
Attempt	Otilia rolled up her wool and, with great skill, began knitting a vest for Frido. “Don’t worry, Frido, I will help you. I’ll knit you a sweater so you won’t be cold.” Meanwhile, Frido sat in front of her and started telling her about the planet he came from. "My planet is called Centaurios, and it’s a place where it’s always warm, so we never wear sweaters.” Otilia replied, "When it's winter on Earth, wearing a sweater is very important because it gets very cold."
Direct Consequence	Frido put on the sweater, and together they came down from the spaceship to play in the meadow. "This is my friend Frido," Otilia told her friends. The other sheep started playing with him.
Reaction and Resolution	Since then, Frido became friends with the sheep and started visiting them every day. The extraterrestrial stayed for the entire winter and never felt lonely again.

APPENDIX 3. Index of Narrative Complexity (Adapted from Bustos & Crespo, 2014).

Narrative Element	0 Points	1 Point	2 Point	3 Point
Character	No main character is included, or only ambiguous pronouns are used. Without the context of an introduction. Example: *She was flying with her friends and was caught by a man and locked in a jar.*	Includes the mention of one or both main characters, but with generic labels. Example: "*the sheep" instead of "Otilia," "the butterfly" instead of "Flopi."*	Includes only one main character with a specific label [specific label and introduction]. *Example: Flopi was caught by a man.*	Includes mention of both main characters with specific labels. Example: *Otilia's body was covered in white wool and she was very cozy.* and *Frido was a green extraterrestrial, with big eyes and a pointed mouth.*
Setting	There is no reference to any specific or general place or time [no markers].	Includes one reference to a general place and time [syntactic marker].	One reference to the specific place and time of the narrative [syntactic and semantic marker].	
Initial Event	No event or problem is mentioned that would likely elicit a response from the character.	Includes at least one event or problem that would likely elicit a response from the character, but there is no direct response related to that event.	Includes at least one event or problem that would likely elicit a response from the character.	Two or more distinct events or problems that elicit a response from the character(s).
Internal Response	There is no mention of the psychological states of the character or characters.	One statement about the psychological state of the character or characters that is not related to any event or problem.	One or more statements about the psychological state of the character causally related to an event or problem.	
Plan	No statement is provided about the character's plan to resolve the event or problem.	A statement is included about how the protagonist or antagonist will resolve the complication or problem they have faced.	Two statements are included about how the protagonist or antagonist will resolve a problem.	Three or more statements about how the character can act or resolve the event(s) or problem(s).
Actions/Attempts	The child does not mention the characters' actions.	Actions carried out by the character are not directly related to the initial event.	The character's actions are oriented toward executing one of the plans in the narrative.	Actions carried out by the character are oriented toward executing more than one of the plans in the narrative.
Consequences	There is no explicit mention of the consequences.	One consequence in one episode.	Two consequences in one episode.	All the consequences in one episode and at least one consequence in a second episode are mentioned.
Formulaic Markers	There are no formulaic markers.	One formulaic marker or one attempt.	Two or more formulaic markers.	
Temporal Markers	There are no temporal markers.	One temporal marker.	Two or more temporal markers.	
Dialogue Knowledge	There is no dialogue.	Dialogue without a speaking character.	Dialogue introduced in a direct or indirect style	
Narrator Evaluations	No evaluation from the narrator.	One evaluation from the narrator.	Two or more evaluations from the narrator.	