Assessing Narrative Competence in Preschool Children: Key Measures and Ensuring Reliability and Validity

BITARA
Volume 8, Issue 1, 2025: 20-43

e-ISSN: 2600-9080
http://www.bitarajournal.com
Received: 8 December 2024
Accepted: 28 December 2024
Published: 16 January 2025

Wu Bo[1] & Jamal Rizal Razali[1*]

1 Centre for Human Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, 26300 Kuantan Pahang, Malaysia
Email: 1052202611@qq.com; jamalrizal@umpsa.edu.my

*Corresponding Author: jamalrizal@umpsa.edu.my

Abstract

This study employed a content analysis approach to review the literature on the assessment of narrative competence in preschool children, focusing on evaluation dimensions, reliability, and validity. The findings reveal that narrative competence assessment primarily revolves around macrostructure and microstructure analyses. Macrostructure analysis emphasizes story grammar and plot coherence, while microstructure analysis examines lexical diversity, syntactic complexity, and productivity. Reliability is predominantly reported through inter-rater reliability metrics, with limited exploration of test-retest reliability or internal consistency. Validity evidence, such as convergent and discriminant validity, highlights the multidimensional nature of narrative competence and its close associations with skills like vocabulary and higher-order cognitive functions. This study underscores the importance of developing reliable and culturally inclusive tools for assessing narrative competence, providing a theoretical foundation for future research and practice in early childhood education.

Keywords: Narrative Competence ; Macrostructure ; Microstructure; Reliability; Validity evidence

Cite This Article:

Wu Bo & Jamal Rizal Razali. (2025). Assessing Narrative Competence in Preschool Children: Key Measures and Ensuring Reliability and Validity. *BITARA International Journal of Civilizational Studies and Human Sciences* 8(1): 20-43.

Introduction

Narrative competence, a key component of early literacy skills, plays a crucial role in child psychology, linguistics, and education. Narrative competence encompasses a child's ability to recount experiences or events in a coherent and logical sequence (Bowles et al., 2020). It is essential for language development, supporting vocabulary expansion, sentence structure acquisition, and the advanced language skills required for reading and writing (Grolig, 2020).

Moreover, narrative skills are strong predictors of future academic performance and play a crucial role in social development (Ralli et al., 2021).

Despite their significance, systematic reviews on the assessment of narrative competence, particularly in preschool children who predominantly produce oral narratives due to limited writing skills, remain scarce (Pinto et al., 2019). The diversity in research subjects, methods, and disciplinary perspectives further underscores the need for a comprehensive analysis to address these gaps and advance the field.

This study aims to conduct a systematic analysis of the literature on preschool children's narrative competence using the WOS database. The study addresses the following research questions, what are the main components and dimensions of assessing preschool children's narrative competence? And how can the reliability and validity of assessments of preschool children's narrative competence be ensured?

Literature Review

Narrative Dimensions

Common types of narratives in children include personal experience narratives and fictional imaginative narratives. Analysing children's narratives can effectively diagnose their language development levels, as it involves linguistic information and examination of language skills such as semantics, grammar, pragmatics, and working memory. Additionally, it evaluates their social understanding abilities (Sun et al., 2024)Currently, there is no standardized approach to studying narrative competence, which allows researchers to explore it from multiple dimensions.

Macrostructure

Macrostructure refers to the underlying structural features in narratives, focusing on the logical relationships within the story (Justice et al., 2006). For children's narratives, the key features of macrostructure include background, characters, and plot. Story grammar and plot complexity are essential indicators when assessing children's narrative abilities. Story grammar typically involves a beginning (setting the scene), middle (initiating event, characters' reactions and plans), and end (resolution of events) (Stein & Glenn, 1979).

Understanding macrostructure helps children comprehend stories more deeply (Kendeou et al., 2009) and serves as a tool for assessing reading comprehension in preschoolers (Paris & Paris, 2001), emphasizing causal relationships and event representations.

Two main models for analysing macrostructure are the climax analysis model and the story grammar model.

The Climax Analysis Model (Labov, 1972) is used for personal experience narratives, focusing on elements like characters, setting, action, viewpoint, resolution, and coda. It categorizes narratives into patterns such as the classic pattern (most complete), ending-at-the-

highpoint, chronological, leapfrogging, and others, depending on the structure and complexity of the events.

Story Grammar Model (Rumelhart, 1975; Thorndyke, 1977) is used for fictional stories, emphasizing the logical consistency of plot and goal-driven problem-solving. A story typically includes background, an initiating event, characters' responses and plans, and the resolution. This model guides children in understanding and organizing the elements of a story, helping them predict the storyline.

Both models focus on essential story components—background, events, and responses—to determine whether a narrative is well-structured and coherent. Story grammar is crucial for creating narratives that can be easily understood, even though more engaging or aesthetically pleasing stories might not always score higher if they lack essential structural elements.

Microstructure

Microstructure analysis focuses on the finer linguistic features within narrative structures. It includes productivity and complexity indicators (Govindarajan & Paradis, 2022). These indicators include Mean Length of Utterance (MLU), Mean Length of the Five Longest Utterances (MLU5), Total Number of Words (TNW), Total Number of Utterances (TNU), Number of Different Words (NDW), Type-Token Ratio (TTR), and Vocabulary Diversity (VOCD).

Language productivity refers to the amount of material generated in a narrative, including TNW and TNU. TNW represents the total number of words that clearly convey meaning in the narrative sample, while TNU assesses the total number of sentences produced in spoken or written language tasks. These two indicators directly reflect the length of the story (Hao et al., 2018).

The Number of Different Words (NDW) is a direct method for measuring the number of different words used in a narrative sample and is a commonly used indicator for assessing vocabulary diversity. Type-Token Ratio (TTR) (TTR = NDW / TNW) is also an indicator of vocabulary diversity but is affected by sample size. To overcome the limitations of TTR, VOCD provides a more stable and reliable measurement of vocabulary diversity (Yang et al., 2022).
Mean Length of Utterance (MLU) is an important indicator for assessing syntactic development and complexity ((Xue et al., 2022)). MLU refers to the average number of morphemes in a child's utterances, calculated by MLU = TNW / TNU. Research has found that the Mean Length of the Five Longest Utterances (MLU5) is a more effective indicator of children's grammatical development.

Micro-level analyses provide insights into how children structure their narratives at both the macro and micro levels, reflecting their language development, comprehension abilities, and overall narrative skills.

In summary, existing research has explored the assessment methods of children's narrative competence from both macrostructure and microstructure perspectives, revealing the

multidimensional characteristics and key indicators of narrative competence. The following will organize and report the specific dimensions and content of children's narrative.

Materials and Methods Section

Search Strategy

A systematic review was conducted using the WOS database, with synonyms incorporated to broaden the search scope without applying expanders. The search targeted key terms related to preschool children, narrative competence, and assessment methods, with a focus on studies published in English. The search syntax included combinations such as: "preschool children" AND "narrative skills" AND ("assessment" OR "evaluation" OR "measurement").

Selection and Inclusion Criteria

After compiling an initial pool of 443 studies, the research entered the screening phase. Studies were included in the final sample only if they met all the specified eligibility criteria. A three-stage filtering process was applied sequentially to refine the sample (see Figure 1 for an overview of the screening process).

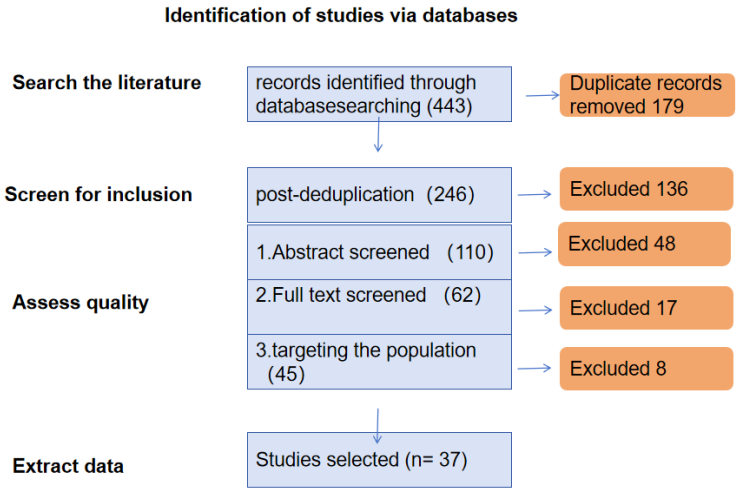

Figure 1. Flow diagram.

The first filter, applied at the abstract level, included studies that met the following criteria: (i) focused on assessing narrative competence in typically developing preschool children, (ii) excluded case studies, (iii) were published between 2019 and 2024, and (iv) were in English (only the manuscript language, not the participants' language background). The second filter, at the full-text stage, further narrowed the selection to studies providing measurement data on oral narratives, excluding those focused on written narratives.

The third filter, targeting the population of interest, included studies with preschool-aged children in the sample, even if other age groups were present, if the focus was on typically developing children. Studies involving children with developmental disabilities or special needs (e.g., language impairment or Down syndrome) or those with Individualized Education Plans (IEPs) were excluded. In total, 37 studies were selected for data extraction.

Data Extraction

The studies were coded and analysed based on three dimensions: evaluation content, reliability, and validity. Table 1 outlines the categories analysed, following the framework by Sanches-Gomez et al. (2024). The coding categories for reliability and validity, as recommended by AERA et al. (2014), included internal consistency, test-retest reliability, inter-rater reliability, content validity, criterion-related validity, convergent validity, and structural validity.

For reliability and validity reporting, relevant or theoretically plausible evidence was included, even if not explicitly stated. For example, studies linking language variables to narrative skill theory were considered validity evidence due to their theoretical relevance.

Two raters independently coded the studies. One rater coded all studies, while the second coded 25% (N=9). Cohen's Kappa (narrative dimensions = 0.90, reliability = 0.88, validity = 0.85) indicated acceptable inter-rater agreement. Discrepancies were resolved through discussion. Table 1 provides a detailed breakdown of the analysis categories.

Table1. Categories of analysis based on the characteristics of the assessment conducted

Criteria	Categories and Examples
Dimensions of analysis	macrostructure (e.g., story grammar); microstructure (e.g., lexical diversity); mixed (e.g., macrostructure and microstructure) .
Type of reliability evidence and reliability data available	e.g., inter-rater reliability (e.g, Krippendorff's alpha, Cohen's kappa); internal consistence (e.g, ordinal alpha, Cronbach's alpha) ; test-retest(e.g., Pearson correlation coefficient)
Sources of validity evidence and validity data available	e.g., Evidence based on test content, internal structure, on relations to other variables (convergent, discriminant, test-criterion relationships).

Results

This section is divided into three subsections. The first analyses the levels of narrative analysis and the content of the analysis in the target documents. The second summarizes the reliability analysis of these data, and the third focuses on the validity of the narrative measurements.

Dimensions and content of analysis

Table 2 provides an overview of the specific criteria for each level of analysis. Overall, these studies effectively tailored their narrative analysis to align with their research objectives.

Table 2. Components or measures analysed and available reliability and validity evidence

Study	Components Analysed	Reliability Data Available	Validity Data Available
Bailey et al. (2020)	Macrostructure : 1.high-point analysis 2.the number of logically sequenced events	Inter-rater reliability, Cohen's kappa for r each dimension	Test-criterion evidence , the criterion- related validity relationship between narrative competence and reading achievement. Convergent validity , Examining the relationship between maternal language support and children's narrative abilities.
Baldwin et al., (2022)	Macrostructure : NSS （story introduction, development, mental and emotional states, referencing/ listener awareness, character conflict/resolution, event/reaction, cohesion and the conclusion）	Inter-rater reliability, Cohen kappa coefficient for each dimension of NSS	
	Microstructure : 1.MLUw 2.NDW	Inter-rater reliability, the percentage of agreement for C-unit segmentation	Convergent Validity: The measurement of narrative competence is consistent with the expected related construct, namely language-based skills.

			Discriminant Validity: There is no significant relationship between narrative competence and AAE dialect density (DDM).
Bohnacker et al. (2022)	Macrostructure：1. story structure (setting, episode, internal state as initiating event, goal, attempt, outcome, internal state as reaction) 2. episodic complexity	unmentioned	Convergent Validity: Narrative length and expressive vocabulary have a significant impact on narrative structure scores.
	Microstructure；1. TNW 2.Expressive vocabulary knowledge		Discriminant Validity: Daily language exposure and the length of Swedish language exposure have no significant impact on narrative structure scores.
Byrnes-Cloet and Hill (2022)	Macrostructure：1. overall text structure 2.coherence Microstructure: word and utterance	unmentioned	
Clark-Whitney and Melzi, （2023)	Macrostructure：1. number of narratives told. 2. narrative organization elements (thematic coherence, setting, elaboration, evaluation, conclusion, temporal and causal markers, character introductions, and anaphora) Microstructure：automated expressive language features (MLUw, VOCD, Unique sophisticated words)	Inter-rater reliability，ICC for Narrative organization elements.	Convergent Validity: The relationship between Executive Function (EF) skills and narrative production abilities. Discriminant Validity: The relationship between EF skills and narrative organization and expressive language skills.

Dealy et al.(2019)	Macrostructural: 1. response to story stem (prosocial aggressive/conflict, avoidance/danger) 2.narrative coherence	Inter-rater reliability, ICC for each dimension	Internal Structure Validity: CFA confirmed a strong link between narrative variables and coherence. Convergent Validity: EF significantly correlates with narrative ability at age 4. Discriminant Validity: Language ability correlates with narrative coherence and prosocial behaviour but not with avoidance/danger or aggressive/conflict behaviours.
Fiani et al. (2022)	Macrostructure1.story structure (setting, initiating event, goal, attempt, outcome, reaction) 2. story structure complexity (total complexity and most complex episode) 3. number of internal state terms		Convergent Validity: The impact of vocabulary ability, age, and language dominance on narrative competence.
Gámez and González (2019)	Macrostructure:1. narrative structure (elements, characters / setting, initiating event, internal response, plan, action / attempt, complication, consequence, and ending)	Inter-rater reliability, Cohen's kappa agreement for story structure	
	Microstructure : 1.story length 2. word diversity（the number of clauses, NDW）	Internal consistency	Convergent Validity: Revisions are positively correlated with vocabulary diversity (NDW). Discriminant Validity: Differences in errors and revisions between children from different language backgrounds
Grolig et	Macrostructure:		

al. (2020)	Narrative comprehension 1.inferential comprehension (dialog, feelings, thoughts, prediction, and theme) 2 literals comprehensions（ characters, setting, initiating event, problem identification, and resolution）	Inter-rater reliability：ICC for Narrative comprehension	
	Microstructure 1. narrative production 2.vocabulary（depth, breadth）	Internal consistency: Cronbach's alpha for some dimension (breadth of vocabulary)	
Guedes et al. (2023)	Macrostructure: elements, sequence, decontextualized	Internal Consistency, Cronbach's alpha for each dimension	Convergent Validity: The study shows a relationship between narrative competence and behavioural self-regulation, supporting convergent validity. Discriminant Validity: The article shows different relationships between narrative competence and behavioural and emotional self-regulation, supporting discriminant validity.
Işıkoğlu and Güzen (2024)	Microstructure: syntax Macrostructure 1. explicit Information (characters， setting, introductory event, problem, result) 2.lmplicit information (emotions, cause/effect ， dialog		Convergent Validity: The relationship between narrative competence and other language abilities.

	, theme)		
Jiménez et al. (2024)	Macrostructure: narrative structured (introduction, characters, a sequence of events, solutions to problems, and a conclusion structure). Microstructure: 1. syntactic complexity of the discourse 2. productivity 3. lexical diversity, 4. accuracy 5. fluency 6. grammaticality	Inter-rater reliability, ICC for some dimensions (narrative structure , unique word measurements, t-unit, free word writing)	Convergent Validity: There is a significant relationship between children's oral narrative competence and writing ability.
Khan, Hong, et al. (2021)	Macrostructure: story structure components (characters, settings, problem, emotional reaction and reason for expressed emotion, attempted solution, and conclusion)	Inter-rater reliability, the percentage of agreement for story structure components Internal Consistency , Cronbach's alpha for narrative comprehension measure	Convergent Validity: Mathematical skills are moderately correlated with narrative comprehension skills. Discriminant Validity: Mathematical language skills uniquely contribute to narrative comprehension.
Khan, Logan, et al. (2021)	1.Macrostructure: identification of main character and identification of subgoals, problems, and resolutions	Inter-rater reliability Cohen's kappa for the item level	
	2Microstructure: use of conjoined adverbial phrases, gratuitous terms, Tier 2 vocabulary, and elaborated noun phrases.	Inter-rater reliability, Cohen's kappa for the item level	Convergent Validity: Lower-level language skills are significantly correlated with narrative ability (NAI). Discriminant Validity: Different levels of narrative ability are associated with different language skills.

Kiernan et al. (2024)	Microstructure: Verb Phrase, followed by Noun Phrase, and Clause Structure	Inter-rater reliability, Percent agreement for Hl, grammatical features and GA	
Lai (2020)	Macrostructure :1. narrative elements (orientation, complicated action, evaluation, resolution, and appendage) 2. narrative patterns	Inter-rater reliability, Cohen's kappa for structural elements and narrative patterns	Construct validity: Narrative skills are directly related to the narrative structure
	Microstructure: 1. linguistic productivity and syntactical complexity 2. evaluative features (Affection Intention, Cognition, Obligation, Sensory, Physiology, Reported, Negation, Gratuitous, Repetition, Word perse)	Inter-rater reliability, Cohen's kappa for some dimensions (coding productive and exceptive language competence, evaluative features)	Construct validity: Narrative skills are directly related to language
Lake and Evangelou (2019)	Macrostructure: story grammar	unmentioned	Convergent validity : The relationship between executive function and narrative test results.
	Microstructure 1. receptive vocabulary 2. productive vocabulary 3. MLU		
Lindgren (2019)	Macrostructure: setting, episode l, is as ie, goal, attempt, outcome, is as r	unmentioned	Convergent Validity: The relationship between narrative ability and other cognitive skills.
Lindgren and Bohnacker (2022)	macrostructure1. story structure（initiating event, goal, attempt, outcome, reaction）2. episodic complexity Microstructure: TNW	Inter-rater reliability, Cronbach's alpha agreement for story structure	Convergent Validity: The correlation between age, language, narrative task, language proficiency, exposure, and narrative macrostructure.

Lindgren (2022)	Macrostructure :1. story comprehension 2.time, place, initiating event, goal, attempt, outcome and internal state	Inter-rater reliability, Cohen's kappa for each dimension	
	Microstructure: productivity, syntactic complexity and referent introduction	Inter-rater reliability, Cohen's kappa for each dimension	
MacLeod and Pesco (2023)	Macrostructure : story grammar (characters, setting, initiating event, internal response, internal plan, attempt, reaction, outcome	unmentioned	Convergent Validity: The relationship between narrative ability and other variables, such as emotional responses and internal states.
Mahfoudhi et al. (2023)	Microstructure 1.productivity (TNW), 2.lexicaldiversity (NDW) 3.syntactic complexity(MLU), 4.the proportion of` complex and compound sen-tences		Internal structure validity: The impact of age and task complexity on narrative microstructure. Convergent validity: The impact of different narrative tasks on the quality of children's language output.
Melzi et al.(2023)	Macrostructure : 1.narrative independence 2.narrative coherence indicators	Inter-rater reliability, Fleiss's kappa for some dimensions(narrative characters and sequences of events)	Convergent Validity: The relationship between the validity of narrative ability and the coherence of temporal sequence and themes.
	Microstructure : 1.syntacticcomplexity of language produced (as measured by MLU-w) 2. length of narrative		
Mendoza et al.	Macrostructural : oral narrative quality	Inter-rater reliability :	

(2021)	Introduction (Resolution, Character Introduction , Narrative Coherence)	Krippendorff's alpha for each dimension Test-retest Test-used to assess the stability of children's oral narrative ability	
	Microstructural : 1.number of conjunctions 2. number of different words 3. communication units 4. mean length of utterance in words 5. mean length of utterance in morphemes	Internal Consistency, Cronbach's alpha for each dimension test-retest Test-used to assess the stability of children's oral narrative ability	Convergent Validity: The correlation between oral narrative quality and microstructure measures.
Orizaba et al. (2020)	Macrostructure: NSS (introduction, character development, mental states, referencing, conflict/resolution, cohesion, conclusion)	Inter-rater reliability, the percentage of agreement in each dimension	Test-criterion evidence : The effectiveness of NAP-S and NSS tools in predicting children's language ability. Construct Validity: The degree of correlation between narrative ability and other language skills.
	Microstructure: NAP-S （sentence structures, phrase structures, modifiers, nouns, and verbs.）	Inter-rater reliability, the percentage of agreement in each dimension	
Pinto et al. (2019)	Macrostructure: 1 structure (title, opening, setting, description of character, description of problem, a central event, resolution to the problem, and story closing) 2 Coherence :the number of incoherencies were counted and balanced for the total number of	Inter-rater reliability percentage for structure and coherence	Convergent Validity: The correlation between different components of narrative ability (structure, coherence, and cohesion). Criterion Validity: The relationship between narrative ability assessment and other related variables.

	propositions		
	Microstructural: cohesion (temporal and causal linguistic connectives were counted, balanced for the total number of words)	Inter-rater reliability percentage for Cohesion	
Pronina et al.(2023)	Macrostructure: story grammar	Inter-rater reliability, percent agreement for overall Inter-rater reliability, Fleiss's kappa for overall	Convergent Validity : The correlation between narrative performance and the accuracy and frequency of gesture use.
Rojas et al. (2019)	Microstructure :1. MLUw; 2. NDW 3. words per minute in English and Spanish		Convergent Validity :Correlation between early literacy skills and language narrative samples
Spencer et al. (2023)	Macrostructure : 1.story grammar 2. inclusion of episodic features	Inter-rater reliability , percentage agreement for each dimension	Convergent Validity: The correlation between narrative ability, executive function (EF), and language ability.
	Microstructure: complex language	Inter-rater reliability , percentage agreement for each dimension	
Sun et al. (2024)	Macrostructure : internal state term, story structure, and structural complexity		Structural validity: The impact of children's Chinese language ability on their narrative ability.
	Microstructure : lexical（ Giraud）and syntactic complexity measures（MLU-W)		Internal structure validity: The correlation between children's Chinese vocabulary-grammar ability and narrative skills
Tompkins et al. (2020)	Macrostructure : 1.Story comprehension 2.Picture sequencing	Inter-rater reliability , the Pearson correlation	Internal structure validity: The internal structural relationship between FBU and narrative

	3.Story generation 4.False Belief Understanding（FBU）	coefficient for some dimensions (story comprehension and story generation)	ability. Convergent validity: The correlation between narrative ability and FBU.
Veneziano et al.(2020)	Microstructure： 1.explanation of events 2. attribution of internal states 3. expression of false belief 4. rectification of false belief 5 markers of causality 6. length of narratives	Cohen kappa coefficient for events explained, internal states, the explanatory function of internal states, the expression of the RFB, Inter-rater reliability, percentage Agreement for causal markers and the number of words in the story.	Internal structure validity: The relationship between narrative complexity and intervention effectiveness. Convergent validity, as it examines the correlation between narrative ability and children's age and developmental stage.
Wofford et al. (2022)	Microstructure: 1.MLU 2.NDW 3. percent utterances with errors		Test-criterion evidence: The correlation between microstructure indicators from narrative samples (e.g., NDW, MLU, PGU) and children's performance in standardized language assessments.
Yang et al. (2022)	Macrostructure： introduction, theme,main character，supporting character(s), conflict, coherence, resolution and conclusion		Convergent Validity: The correlation between age, language experience, and narrative ability.
	Microstructure: 1.general microstructures (TNW,NDW, TNC,MLUw)	Inter-rater reliability, Cohen's kappa for some dimensions (fine-grained	

	2. fine-grained microstructures (phrase structure, modifier, nominal, and verb)	microstructure)	
Yang et al. (2023)	Macrostructure：1.SS（initiating event, goal, attempt, outcome）2.SC (story complexity) 3. IS (internal state terms)		Convergent Validity: The correlation between external factors and narrative ability.
Zanchi and Zampini (2021)	Macrostructure : quantity of information , structure, mental state lexicon	Inter-rater reliability Percent agreement for each dimension	Test-criterion evidence: Children's narrative ability is related to their language and cognitive abilities. Content validity: The NCT test comprehensively assesses key aspects of children's narrative ability.
	Macrostructure : lexical diversity and MLU		
Zhang et al. (2019)	Macrostructure : 1.narrative components, complicating action (specific events), orientation, evaluation, resolution , appendages direct speech, reported speech 2.narrative patterns Microstructure: basic language productivity （TNU, MLU, TNW, NDW）	Inter-rater reliability , the percentage of agreement in Narrative Components and Narrative Patterns	

Note1 : MLU =Mean Length of Utterance; MLUw= Mean Length of Utterance in Words; NDW =Number of Different Words; TNW= Total Number of Words; VOCD =Vocabulary Diversity; TNC =Total Number of Clauses

Most studies (N=21) analyzed both macrostructure and microstructure, while 11 focused only on macrostructure and 5 on microstructure. Macrostructure analysis focuses on key elements of children's narratives, including story grammar (e.g., characters, setting, events, resolution), narrative organization (e.g., thematic coherence, causal markers), and narrative complexity (e.g., plot complexity, emotional expression). Story grammar is a common method for assessing narrative completeness. Some studies focus on specific components, such as "internal response" and "outcome" (MacLeod & Pesco, 2023), or "evaluation" and "resolution" (Zhang et al., 2019), as these elements define what constitutes a "good" story.

The analysis of narrative organizational elements helps determine the clarity, coherence, and expressiveness of the story, reflecting children's cognitive ability to understand and organize events logically (Grolig, Lindgren, & Tompkins). Plot complexity analysis evaluates multi-layered plots, event coherence, and the integration of emotional responses, providing insights into children's abilities in linguistic organization, event construction, and logical reasoning (Pinto.2019).

Microstructure analysis is more variable and includes dimensions such as language length (e.g., total words, mean length of utterance), lexical diversity (e.g., use of unique vocabulary), syntactic complexity (e.g., proportions of compound and complex sentences), and grammatical accuracy. Some studies also focus on the expression of complex linguistic features (e.g., connectives, modifiers) and the use of rhetorical and emotional vocabulary, revealing both the quantitative and expressive aspects of language production.

Reliability Evidence Reported

Reliability refers to the degree to which scores from an instrument are stable and consistent, regardless of how this consistency is estimated or reported (Creswell, 2012). This report focuses on analysing the reliability data for measuring narrative competence across various studies (see Table 2).

While some studies include reliability data for other aspects, this analysis focuses on narrative competence. The studies are categorized into three types for a more precise analysis:

Comprehensive Reliability Data (N=10): Studies providing reliability data for both macrostructure and microstructure Partial Reliability Data: Studies reporting reliability for specific components, further divided into:

i. Narrative-level reliability only (N=9): Focused on macrostructure or microstructure without considering transcription or coding.
ii. Process-level reliability only (N=7): Focused on transcription or coding reliability y without analysing narrative-level reliability.
iii. Combined Narrative and Process Reliability:(N=6): Addressing narrative-level and transcription/coding reliability.
iv. No Reliability Data: Studies that provide no reliability data (N=5).

Transcription and coding are essential for both macrostructure and microstructure assessments (Sánchez-Gómez et al., 2024). Transcription converts oral language to written form, while coding involves classifying and scoring various aspects, such as narrative structure, linguistic accuracy, and syntactic complexity However, such reliability assessments are not directly categorized under the reliability analyses of macrostructure or microstructure. Reliability in these processes is crucial for reliability in measuring narrative competence, but they are categorized as partial reliability data in this report.

Reliability types include inter-rater reliability, internal consistency, and test-retest reliability. Some studies report multiple types. Most studies focus on inter-rater reliability (N=30), which includes transcription and coding reliability. Studies reporting transcription reliability (N=11), coding reliability (N=10); or both (N=6) are analysed separately.

Some studies also report internal consistency (e.g.,Gámez and González, 2019; Grolig et al., 2020; Guedes et al., 2023; Khan, Hong, et al., 2021;Mendoza et al., 2021) which checks if measurement items are interrelated, ensuring the instrument's validity and reliability. A few studies report test-retest reliability (e.g., Mendoza et al., 2021;Grolig et al., 2020), assessing result stability over time.

Reliability coefficients used for these types include Cohen's Kappa for two raters, Fleiss's Kappa for three or more raters, and ICC (Intraclass Correlation Coefficient) for quantitative data. For internal consistency, Cronbach's Alpha is the most common, though Split-half reliability and Omega Coefficient are alternatives. Test-retest reliability often uses Pearson Correlation or the test-retest reliability coefficient, depending on data and objectives.

Validity Evidence Reported

Validity refers to the extent to which evidence and theory support test score interpretations for their intended purposes (AERA et al., 2014). This report analyzes validity evidence from various sources, detailed in Table 2, with results outlined in Table 6, focusing on the relationship between narrative competence and other variables (e.g., convergent, discriminant, and criterion-related validity).

Convergent validity (N=27) examines relationships between narrative competence and various constructs, including language skills (vocabulary, grammar, semantics, phonological awareness), cognitive abilities (executive function), emotional regulation, and educational outcomes (reading, writing, and math). These findings validate narrative competence as a multidimensional ability.

Discriminant validity (N=8) demonstrates that narrative ability is distinct from other constructs. Studies show no significant relationships with factors such as dialect density, language exposure, executive function, or behavior patterns (e.g., aggression or avoidance). Additionally, narrative ability differs by language background and is independent of age and developmental stages. These results support the independence of narrative competence.

Test-criterion evidence (N=4) links narrative competence to reading achievement, False Belief Understanding (FBU), and variables like nonverbal intelligence and reasoning (Zanchi

& Zampini, 2021). Correlations between microstructure indicators (e.g., NDW, MLU, PGU) and standardized language assessments further reinforce the external validity of narrative competence tools.

This analysis highlights the multidimensional nature of narrative competence and its relationship with other key variables, providing a solid foundation for educational and psychological assessments. It also enhances the confidence in narrative competence measurement tools and offers guidance for future research.

Discussion

The assessment of narrative competence in preschool children lacks a unified standard, with researchers focusing on different aspects of children's narrative abilities. While most studies emphasize story grammar, narrative organization, and complexity, other dimensions are occasionally considered. Macrostructure analysis is relatively consistent, while microstructure assessment shows greater variability.

In terms of reliability, while some studies report internal consistency (N=5) and test-retest reliability (N=2), most (N=30) focus on inter-rater reliability, including transcription and coding. However, the reliability of macrostructure and microstructure analyses remains underexplored. Some studies also lack population-specific reliability and validity evidence, raising concerns about the generalizability of tools.

Narrative ability is recognized as a multidimensional construct linked to linguistic skills, cognitive abilities, and socio-emotional regulation, influenced by factors such as age, language exposure, task type, and educational achievements. While some studies provide strong validity evidence, areas like predictive validity and cross-cultural considerations remain underexplored. Some studies omit reliability and validity data, which may not reflect lower quality but indicate that tool development was not their focus. This points to a gap in measurement practices (Sánchez-Gómez et al., 2024). These findings highlight the need for further research on the reliability and validity of narrative tools in different contexts, including cross-cultural assessments and less explored validity types.

Conclusion

The systematic review reveals the complexity and multidimensional nature of assessing narrative competence in preschool children. Macrostructure analysis, focusing on story grammar and thematic coherence, offers insights into children’s ability to organize and structure narratives. Microstructure analysis delves into linguistic features like lexical diversity and syntactic complexity, shedding light on their linguistic development. While inter-rater reliability is commonly addressed, gaps in test-retest and internal consistency raise concerns about the robustness of certain tools. Validity evidence underscores the strong relationship between narrative competence and linguistic, cognitive, and socio-emotional skills, yet predictive validity and cultural adaptability remain underexplored. The findings highlight the

need for culturally inclusive tools to ensure accurate and meaningful assessments. This study lays the groundwork for future research to bridge these gaps, ultimately enhancing the efficacy of narrative competence evaluations in early education and cross-cultural contexts.

References

Bailey, A. L., Moughamian, A. C., Kelly, K. R., McCabe, A., & Huang, B. H. (2020). Leap-frog to literacy: maternal narrative supports differentially relate to child oral language and later reading outcomes. *Early Child Development and Care*, 190(7), 1136–1149. https://doi.org/10.1080/03004430.2018.1521807

Baldwin, E., Heilmann, J., Finneran, D., Cho, C. C., & Moyle, M. (2022). Dialect density, language abilities and emergent literacy skills of prekindergarten children who speak African American English. *Journal of Research in Reading*, 45(4), 567–586. https://doi.org/10.1111/1467-9817.12410

Bohnacker, U., Lindgren, J., & Öztekin, B. (2022). Storytelling in bilingual Turkish-Swedish children. *Linguistic Approaches to Bilingualism*, 12(4), 413–445. https://doi.org/10.1075/lab.20057.boh

Bowles, R. P., Justice, L. M., Khan, K. S., Piasta, S. B., Skibbe, L. E., & Foster, T. D. (2020). *Development of the Narrative Assessment Protocol-2: A Tool for Examining Young Children's Narrative Skill*. https://doi.org/10.23641/asha

Byrnes-Cloet, H., & Hill, S. (2022). Writing Play-Books with Linguistically Diverse Young Learners. *Reading Teacher*, 75(4), 413–427. https://doi.org/10.1002/trtr.2056

Creswell, J. W. (2012). *Education Research Planning Conducting and Evaluating Quantitative and Qualitative Research* (fourth). Pearson.

Dealy, J., Mudrick, H., & Robinson, J. (2019). Children's narrative story stem responses: Contributions of executive functioning and language proficiency to relationship representations. *Social Development*, 28(1), 168–185. https://doi.org/10.1111/sode.12335

Fiani, R., Henry, G., & Prévost, P. (2022). Macrostructure in narratives produced by Lebanese Arabic-French bilingual children. *Linguistic Approaches to Bilingualism*, 12(4), 446–478. https://doi.org/10.1075/lab.20035.fia

Gámez, P. B., & González, D. (2019a). A comparison of narrative skill in Spanish-English bilinguals and their functionally monolingual Spanish-speaking and English-only peers. *International Journal of Bilingualism*, 23(1), 329–348. https://doi.org/10.1177/1367006917728391

Gámez, P. B., & González, D. (2019b). A comparison of narrative skill in Spanish-English bilinguals and their functionally monolingual Spanish-speaking and English-only peers. *International Journal of Bilingualism*, 23(1), 329–348. https://doi.org/10.1177/1367006917728391

Govindarajan, K., & Paradis, J. (2022). Narrative macrostructure and microstructure profiles of bilingual children with autism spectrum disorder: differentiation from bilingual children with developmental language disorder and typical development. *Applied Psycholinguistics*, 43(6), 1359–1390. https://doi.org/10.1017/S0142716422000406

Grolig, L. (2020). Shared Storybook Reading and Oral Language Development: A Bioecological Perspective. *Frontiers in Psychology*, 11. https://doi.org/10.3389/fpsyg.2020.01818

Grolig, L., Cohrdes, C., Tiffin-Richards, S. P., & Schroeder, S. (2020). Narrative dialogic reading with wordless picture books: A cluster-randomized intervention study. *Early Childhood Research Quarterly*, 51, 191–203. https://doi.org/10.1016/j.ecresq.2019.11.002

Guedes, C., Ferreira, T., Leal, T., & Cadima, J. (2023). Unique and joint contributions of behavioral and emotional self-regulation to school readiness. *Applied Developmental Science*, 27(2), 136–155. https://doi.org/10.1080/10888691.2022.2045200

Hao, Y., Sheng, L., Zhang, Y., Jiang, F., De Villiers, J., Lee, W., & Liu, X. L. (2018). A narrative evaluation of mandarin-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, 61(2), 345–359. https://doi.org/10.1044/2017_JSLHR-L-16-0367

Işıkoğlu, N., & Güzen, M. (2024). The promise of digital storytelling for kindergarteners: language and technology skills. *Early Child Development and Care*, 194(2), 195–207. https://doi.org/10.1080/03004430.2023.2299389

Jiménez, J. E., Rodríguez, C., & Balade, J. (2024). Identifying kindergarteners at-risk of writing difficulties based on foundational literacy skills. *Reading and Writing*. https://doi.org/10.1007/s11145-024-10518-7

Khan, K. S., Hong, F., Justice, L. M., Sun, J., & Mills, A. K. (2021). Cross-domain associations between mathematical and narrative abilities in preschool-aged children. *Journal of Experimental Child Psychology*, 212. https://doi.org/10.1016/j.jecp.2021.105233

Khan, K. S., Logan, J., Justice, L. M., Bowles, R. P., & Piasta, S. B. (2021). The contribution of vocabulary, grammar, and phonological awareness across a continuum of narrative ability levels in young children. *Journal of Speech, Language, and Hearing Research*, 64(9), 3489–3503. https://doi.org/10.1044/2021_JSLHR-20-00403

Kiernan, R., Pearce, W., & Flanagan, K. (2024). Home language variation in the narratives of urban First Nations Australian children in their first year of school. *Clinical Linguistics and Phonetics*, 38(7), 626–641. https://doi.org/10.1080/02699206.2023.2233048

Lai, W. F. (2020). Narrative Structure and Evaluative Language of Personal Narratives Shared by Urban Aboriginal Children in Taiwan. *Early Child Development and Care*, 190(14), 2214–2228. https://doi.org/10.1080/03004430.2019.1566229

Lake, G., & Evangelou, M. (2019). Let's Talk! An interactive intervention to support children's language development. *European Early Childhood Education Research Journal*, 27(2), 221–240. https://doi.org/10.1080/1350293X.2019.1579549

Lindgren, J. (2019). Comprehension and production of narrative macrostructure in Swedish: A longitudinal study from age 4 to 7. *First Language*, 39(4), 412–432. https://doi.org/10.1177/0142723719844089

Lindgren, J. (2022). The development of narrative skills in monolingual Swedish-speaking children aged 4 to 9: a longitudinal study. *Journal of Child Language*, 49(6), 1281–1294. https://doi.org/10.1017/S030500092100057X

Lindgren, J., & Bohnacker, U. (2022). How do age, language, narrative task, language proficiency and exposure affect narrative macrostructure in German-Swedish bilingual children aged 4 to 6? *Linguistic Approaches to Bilingualism*, 12(4), 479–508. https://doi.org/10.1075/lab.20020.lin

MacLeod, A. A. N., & Pesco, D. (2023). Narratives by bilingual children: a tale of strengths and growth during kindergarten. *International Journal of Bilingual Education and Bilingualism*, 26(7), 891–903. https://doi.org/10.1080/13670050.2022.2132107

Mahfoudhi, A., Abdalla, F., & Al-Sulaihim, N. (2023). The Effect of Age and Task Complexity on the Microstructure of Child Arabic Narratives. *Language, Speech, and Hearing Services in Schools*, 54(3), 952–966. https://doi.org/10.1044/2023_LSHSS-22-00091

Melzi, G., Schick, A. R., & Wuest, C. (2023). Stories beyond Books: Teacher Storytelling Supports Children's Literacy Skills. *Early Education and Development*, 34(2), 485–505. https://doi.org/10.1080/10409289.2021.2024749

Mendoza, M., Beltran-Navarro, B., Matute, E., & Rosselli, M. (2021a). Effects of the age, sex, and maternal education of monolingual spanish-speaking preschool children on oral narrative production. *Journal of Speech, Language, and Hearing Research*, 64(2), 579–592. https://doi.org/10.1044/2020_JSLHR-19-00175

Mendoza, M., Beltran-Navarro, B., Matute, E., & Rosselli, M. (2021b). Effects of the age, sex, and maternal education of monolingual spanish-speaking preschool children on oral narrative production. *Journal of Speech, Language, and Hearing Research*, 64(2), 579–592. https://doi.org/10.1044/2020_JSLHR-19-00175

Orizaba, L., Gorman, B. K., Fiestas, C. E., Bingham, G. E., & Terry, N. P. (2020). Examination of narrative language at microstructural and macrostructural levels in spanish-speaking preschoolers. *Language, Speech, and Hearing Services in Schools*, 51(2), 428–440. https://doi.org/10.1044/2019_LSHSS-19-00103

Pinto, G., Tarchi, C., & Bigozzi, L. (2019a). Promoting narrative competence in kindergarten: An intervention study. *Early Childhood Research Quarterly*, 47, 20–29. https://doi.org/10.1016/j.ecresq.2018.09.003

Pinto, G., Tarchi, C., & Bigozzi, L. (2019b). Promoting narrative competence in kindergarten: An intervention study. *Early Childhood Research Quarterly*, 47, 20–29. https://doi.org/10.1016/j.ecresq.2018.09.003

Pronina, M., Grofulovic, J., Castillo, E., Prieto, P., & Igualada, A. (2023). Narrative Abilities at Age 3 Are Associated Positively With Gesture Accuracy but Negatively With Gesture Rate. *Journal of Speech, Language, and Hearing Research*, 66(3), 951–965. https://doi.org/10.1044/2022_JSLHR-21-00414

Ralli, A. M., Kazali, E., Kanellou, M., Mouzaki, A., Antoniou, F., Diamanti, V., & Papaioannou, S. (2021). Oral Language and Story Retelling During Preschool and Primary School Years: Developmental Patterns and Interrelationships. *Journal of Psycholinguistic Research*, 50(5), 949–965. https://doi.org/10.1007/s10936-021-09758-3

Rojas, R., Hiebert, L., Gusewski, S., & Francis, D. J. (2019). Moving Forward by Looking Back: Understanding Why Some Spanish-Speaking English Learners Fall Behind. In *New Directions for Child and Adolescent Development* (Vol. 2019, Issue 166, pp. 43–77). Wiley-Blackwell Publishing. https://doi.org/10.1002/cad.20305

Sánchez-Gómez, V., Verdugo, M. Á., Calvo, M. I., Amor, A. M., Palomero-Sierra, B., & Zampini, L. (2024). How to Assess Oral Narrative Skills of Children and Adolescents with Intellectual Disabilities: A Systematic Review. *Behavioral Sciences*, 14(4), 308. https://doi.org/10.3390/bs14040308

Spencer, T. D., Kirby, M. S., Garcia, A. R., & Sturm, B. (2023). Additive Effect, Feasibility, and Cost-Effectiveness of Family Engagement Activities that Supplement Dual Language Preschool Instruction. *Early Childhood Education Journal*. https://doi.org/10.1007/s10643-023-01571-y

Sun, H., Tan, J., & Feng, L. (2024). Proficiency and compensatory strategies in bilingual children's Mandarin Chinese narrative. Reading and Writing, 37(7), 1733–1752. https://doi.org/10.1007/s11145-024-10541-8

Tompkins, V., Blosser, M. K., & Downing, M. (2020). False belief understanding and narrative comprehension in the preschool years. *Cognitive Development*, 56. https://doi.org/10.1016/j.cogdev.2020.100936

Veneziano, E., Le Normand, M.-T., Plumet, M.-H., & Elie-Deschamps, J. (2020). Promoting narrative skills in 5-to 8-year-old French-speaking children: The effects of a short conversational intervention. *First Language*, 40(3), 225–250. https://doi.org/10.1177/0142723720901614

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. In *Language, Speech, and Hearing Services in Schools* (Vol. 53, Issue 2, pp. 511–531). American Speech-Language-Hearing Association. https://doi.org/10.1044/2021_LSHSS-21-00055

Xue, J., Zhuo, J., Li, P., Liu, J., & Zhao, J. (2022). Characterizing macro- and micro-structures of narrative skills for Mandarin-speaking school-age children with specific language impairment. *Journal of Communication Disorders*, 96, 106199. https://doi.org/10.1016/j.jcomdis.2022.106199

Yang, J., Kim, J. H., Tuomainen, O., & Xu Rattanasone, N. (2022). Bilingual Mandarin-English preschoolers' spoken narrative skills and contributing factors: A remote online story-retell study. *Frontiers in Psychology*, 13. https://doi.org/10.3389/fpsyg.2022.797602

Yang, J. S., Rosvold, C., & Bernstein Ratner, N. (2022). Measurement of Lexical Diversity in Children's Spoken Language: Computational and Conceptual Considerations. *Frontiers in Psychology*, 13. https://doi.org/10.3389/fpsyg.2022.905789

Yang, W., Chan, A., & Gagarina, N. (2023). Left-behind experience and language proficiency predict narrative abilities in the home language of Kam-speaking minority children in China. *Frontiers in Psychology*, 13. https://doi.org/10.3389/fpsyg.2022.1059895

Zanchi, P., & Zampini, L. (2021). The narrative competence task: A standardized test to assess children's narrative skills. *European Journal of Psychological Assessment*, 37(1), 15–22. https://doi.org/10.1027/1015-5759/a000569

Zhang, F., McCabe, A., Ye, J., Wang, Y., & Li, X. (2019). A Developmental Study of the Narrative Components and Patterns of Chinese Children Aged 3–6 Years. *Journal of Psycholinguistic Research*, 48(2), 477–500. https://doi.org/10.1007/s10936-018-9614-3