frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 08 August 2022
DOI 10.3389/feduc.2022.918127

OPEN ACCESS

EDITED BY
Sarah Powell,
University of Texas at Austin,
United States

REVIEWED BY
Lénia Carvalhais,
Infante D. Henrique Portucalense University, Portugal
Ingrida Balèiûnienë,
Vytautas Magnus University, Lithuania

*CORRESPONDENCE
Sandra L. Gillam
sandi.gillam@usu.edu

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

RECEIVED 12 April 2022
ACCEPTED 21 July 2022
PUBLISHED 08 August 2022

CITATION
Israelsen-Augenstein M, Fox C, Gillam SL, Holbrook S and Gillam R (2022) Monitoring indicators of scholarly language: A progress monitoring tool for documenting changes in narrative complexity over time.
Front. Educ. 7:918127.
doi: 10.3389/feduc.2022.918127

Monitoring indicators of scholarly language: A progress monitoring tool for documenting changes in narrative complexity over time

Megan Israelsen-Augenstein[1], Carly Fox[2], Sandra L. Gillam[2]*, Sarai Holbrook[3] and Ronald Gillam[2]

[1]Communication Sciences and Disorders Division, West Virginia University, Morgantown, WV, United States, [2]Communicative Disorders and Deaf Education, Utah State University, Logan, UT, United States, [3]Communication Sciences and Disorders, University of Wisconsin – Stevens Point, Stevens Point, WI, United States

The purpose of this cross-sectional study was to determine the differences in narrative macrostructure abilities of children in different age groups using a progress monitoring tool based in discourse theory. A majority of existing research regarding narrative developmental patterns has been based in schema theory. The *Monitoring Indicators of Scholarly Language* (MISL) rubric is based in discourse theory and was designed to characterize aspects of narrative proficiency in school-age children. The data for this project consisted of 687 narratives elicited using the Aliens subtest from The *Test of Narrative Language—Second Edition* (TNL-2). There were 1,597 participants who ranged in age from 4; 0 to 15; 0 (year; month). An ordinary least squares regression where age predicted total macrostructure score, followed by a series of *post hoc* ordinal logistic regressions (OLR) where age predicted each individual MISL rubric element was used. Results of both the simple regression on total macrostructure score and the series of ordinal regression analyses for each macrostructure element indicated that age was a significant predictor of the scores children received. Collectively, these results suggest that the MISL is a developmentally valid measure of narrative production abilities. Developmental milestones based on discourse theory are reported to be substantially later than has been reported for schema theory. The differences are highlighted and the implications for progress monitoring for narrative development are discussed.

KEYWORDS

narrative language, progress-monitoring measures, narrative macrostructure, language impairment, narrative discourse

Introduction

The study of narrative discourse is a critical pursuit in the field of speech language pathology, particularly for professionals who work with school-aged children. Discourse can be defined as text or spoken language beyond the sentence level (Hughes et al., 1997; Nicolosi et al., 2004), while narratives are a genre of discourse also known as stories (Berman and Nir-Sagiv, 2007; Graham et al., 2013; Dockrell et al., 2014). Knowledge and use of narrative discourse requires a child to produce stories that contain specific structural features of narrative language and serves a specific communicative goal (Berman and Nir-Sagiv, 2007; Carvalhais et al., 2021). Narrative discourse is valued in the study of school-aged children's language because the ability to successfully produce a narrative is considered an important developmental milestone and is included in the Common Core State Standards for students in the United States (National Governors Association Center for Best Practices and Council of Chief State School Officers, 2010). In addition, research has suggested that preschool and school-age children who struggle with narrative production and comprehension are more likely to experience later academic difficulties in tasks involving reading, writing, and oral language (Liles et al., 1995; Greenhalgh and Strong, 2001; Catts et al., 2002; Roth et al., 2002; Justice et al., 2006; Gillam et al., 2017). Narrative comprehension and production require a complex integration of social, linguistic, pragmatic, and cognitive skills that make it an ideal method for studying a child's communication abilities (Liles, 1993; Wagner et al., 2000; Botting, 2002; Nippold et al., 2014). Due to their complex nature, narratives can and are used as a measure of language ability for students in a wide age range (MacLachlan and Chapman, 1988; Dollaghan et al., 1990; Leadholm and Miller, 1992; Wagner et al., 2000; Westerveld et al., 2004; Nippold et al., 2014). While most researchers agree that typically developing children produce "adult-like" narratives by the age of six or seven (Hughes et al., 1997), there is evidence that narratives continue to grow in complexity through adolescence (Applebee, 1978; Peterson and McCabe, 1983; Roth and Speckman, 1986; Purcell and Liles, 1992; Liles, 1993; Crais and Lorch, 1994; Munoz et al., 2003; Stadler and Ward, 2005).

This makes the evaluation of narrative discourse skill, which is often conducted through language sample analysis, a unique context in which to gain a more complete picture of a child's language profile over time. Therefore, a number of progress monitoring tools have been designed to make discourse level analysis more accessible to speech language pathologists. These tools have largely been based in schema theory, which has been the prevailing conceptualization of narrative discourse structure used in the field of speech language pathology and education. In schema theory, narratives are largely defined by the elements they contain (Meyer, 1975; Mandler and Johnson, 1977; Rumelhart, 1977; Thorndyke, 1977; Stein and Glenn, 1979) and consist of a setting and episode (Stein and Glenn, 1979). The setting system consists of the main character(s) and the physical, temporal and/or spatial location of the story. An episode includes the main character, a goal, actions/attempts directed at achieving that goal (often referred to as attempts), and a consequence or resolution. More complex episodes include the main character's internal responses (feelings) related to the goal, and plan(s) to achieve their goal. Aspects of the story (character, setting, goals, plans, actions, consequences) are referred to as story grammar elements (SGEs).

Schema theory views a child's ability to comprehend or produce a narrative as related to their internal organization or knowledge of story grammar elements (Rumelhart, 1975; Mandler and Johnson, 1977; Stein and Glenn, 1979). Schemata, in a sense, form our expectations of what components a story should possess, so that information can be processed more efficiently. Stein and Glenn (1979) proposed developmental stages for the types of narratives that children produce with preschoolers often telling stories that "describe characters" and "list actions" in a temporal order. At about 6 years of age, children are said to tell stories that include the aims or intentions of the character but may not include specific names of their characters. Between the ages of 7–8, children were reported to begin telling stories that have a "chain of reactive sequences" or "abbreviated episodes." Key elements included in an abbreviated episode include an initiating event with a chain of actions taken by the characters. Stories that include an abbreviated episode often omit a conclusion.

A full episode, according to Stein and Glenn (1979), contains a basic episode and is produced by children 8–9 years of age. A story that includes a full episode includes an initiating event, attempts, and a consequence that are related and included in a cohesive and sequential order. By the time a child is 11, they are said to tell stories that are complex and include elaborated episodes with multiply embedded plans, and/or attempts.

A longitudinal study conducted by Berman and Nir-Sagiv (2007) is one of the only studies that has been conducted to substantiate the developmental stages proposed by Stein and Glenn (1979). The researchers analyzed written texts produced by 80 English-speaking children and adults. The participants were split into four age groups (i.e., elementary school children, junior high school students, high school students, and university students). Each participant was asked to write a narrative retell from a video prompt. These stories were then coded for different linguistic and narrative elements to determine the complexity of the narratives produced by individuals in each age group. They reported that story retells produced by the elementary school children in this study differed from reports of stories produced by preschool children. Preschool children typically produce stories that often contain weakly developed narrative macro- and microstructure elements. For example, the children do not produce a wide variety of macrostructure elements (i.e., story grammar elements; SGEs), and fail to include microstructure elements like subordinating and coordinating conjunctions. The

researchers proposed that the preschool children in their studies may not have had the linguistic and cognitive skills necessary to establish causal and temporal relationships within their narratives. However, by early school-age (around 4th grade, or approximately ages 8–9), they determined that the children were developing linguistic foundations that allowed them to include basic story elements in their written stories. These children did not yet produce narratives that were elaborate and included clear causal connections. The data gathered from this study indicated that children may develop the ability to produce well-formed narratives that utilize all SGEs after fourth grade. The researchers' findings differed from Stein and Glenn (1979) because it wasn't until children were older (i.e., junior high school students) that they produced written narratives that were well developed and included complex and elaborated episodes. These findings suggest that oral and written narrative development may differ slightly, however, similar development patterns are observed. Knowing the trajectory of narrative development can assist speech-language pathologists in applying and understanding the needs of children when using progress-monitoring tools to assess narrative abilities of children on their caseloads.

Assessment tools based in schema theory require the examiner to note the presence or absence of specific story elements used by the storyteller and have been elicited from a range of story prompts (e.g., story retells, sequenced pictures). Children who include certain elements, or "more" story elements are thought to have better narrative abilities than children who omit important story elements or use fewer elements (Berman, 1988; Strong, 1998; Boudreau and Hedberg, 1999; Miles and Chapman, 2002; Reilly et al., 2004). Other scoring rubrics incorporate subjective "text-level" judgments to rate overall story quality (Applebee, 1978; Stein, 1988; Hedberg and Westby, 1993).

For example, The Strong Narrative Assessment Procedure (SNAP; Strong, 1998) is a tool aligned with schema theory designed to measure both macro- and microstructure elements of narratives. Standardized samples of text-level discourse are elicited using audiotaped stories that narrate the wordless pictures books: A Boy, a Dog and a Frog (Mayer, 1967), Frog, Where Are You? (Mayer, 1969), Frog Goes to Dinner (Mayer, 1974), and One Frog Too Many (Mayer, 1975). Children are asked to retell each of these stories after listening to them. The story retells are recorded, transcribed and analyzed for 26 different narrative macrostructure and microstructure elements. This assessment provides information about overall use of SGEs and general language features in the stories children produce. Both of these assessments provide information related to the knowledge and use of specific types of SGEs used by children in their stories. Differences in narrative performance across ages have been documented in narrative retell tasks using this assessment and are comparable to the developmental data reported by Stein and Glenn (1979) and Berman and Nir-Sagiv (2007) (see Strong, 1998; John et al., 2003).

John et al. (2003) completed a study to determine if the SNAP yielded differences in story retelling abilities in children across different ages. Story retell samples were elicited from 61 typically developing children between the ages of 6 and 11. The children were assigned to three different age groups for the purpose of data analysis. The SNAP assessment (i.e., story grammar analysis) was used to score the story retell samples that were elicited using four wordless picture books created for the study. The researchers found that the mean scores for the proportion of story grammar elements retold were consistent with previous literature (e.g., Stein and Glenn, 1979) because the children in their sample recalled initiating events, attempts, and consequences more often than elements like internal response. Age was found to be a significant predictor for the story grammar element of internal response. The children in the youngest age group recalled significantly fewer instances of internal response as compared to the other two age groups. In addition, the children in the oldest age group (i.e., 11-year-olds) reported the element of internal response more often than those in either of the other groups. Children in this sample were demonstrating the use and recall of elements included in a basic story episode (i.e., initiating event, attempt, and consequence), as well as recalling instances of internal response by the time they were 7 years old. As children grew older, they demonstrated a greater number of recalls of instances of internal response, with children who were 11 years old including a significantly larger amount than all other children included in the study. Using the SNAP, a story grammar analysis is gathered that has been shown to yield differences in narrative story retelling ability across age. However, information regarding the causal and temporal connections established by children in stories and the global organization and coherence of a story is not gleaned using this assessment. This information may be necessary in order to more completely understand a child's narrative discourse abilities.

An alternate understanding of narrative ability comes from discourse theory, which is a broad subfield of linguistics dedicated to the study of language and communication beyond the sentence level. Here, we use the term discourse theory to refer to the narrower scope of discourse processing, as described in the Event-Indexing model, which focuses on the construction of a mental representation during narrative comprehension and production (Zwaan et al., 1995). Discourse theory can be thought of as an extension of schema theory that aims to account for a larger number of variables. Discourse theorists recognize the importance of schemata in narrative production and comprehension, as schemata as it allows for the child to more efficiently and accurately recall story information and adapt/monitor their mental representation of the story content (Kintsch and van Dijk, 1978; van Dijk and Kintsch, 1983; Zwaan et al., 1995). However, discourse theory provides a more in-depth explanation for how local and global coherence in narratives (or more generally discourse) are established. Kintsch and van Dijk (1978) highlighted the importance of

forming a coherent text-base in order to create a fully formed situation (mental) model of the narrative. They proposed that when processing a narrative, the individual clauses are reviewed at the proposition level (i.e., a predicate and argument) and are compared for argument overlap within working memory. From this point of view, the likelihood that a proposition is stored and subsequently influences the mental representation of the narrative increases as a function of the amount of times they overlap with other propositions across causal (i.e., the causal connections between events) and temporal (i.e., the temporal relationship between events) dimensions (Zwaan et al., 1995). The greater the overlap, the more likely a particular proposition is to be important to the central theme or plot. The explicit inclusion of causal and temporal connections between events in a story is a critical component of evaluation for discourse theory-based narrative assessment tools, as it provides an objective measure of the storyteller's understanding of the relationship between events in a story. Therefore, narrative progress monitoring tools based in discourse theory require a measurement of the level and overlap of propositions across causal and temporal events in a story in addition to a measurement of the use and knowledge of SGEs. Notably such measures of causal and temporal events in a story are not seen in tools designed from schema -theory.

One of the first tools to incorporate both aspects (discreet scores for story elements and holistic ratings of overall story quality) was the Narrative Scoring Scheme (NSS; Miller et al., 2003; Heilmann et al., 2010). The Narrative Scoring Scheme was developed to measure the use of specific story elements as well as overall story "quality" using story retelling of Frog Stories (e.g., Mayer, 1967, 1969, 1974, 1975). Key story elements measured include an introduction, conflicts, and the conclusion of the story. These elements constitute the "macrostructure" analysis, whereby other aspects of language microstructure are measured by noting the presence or absence of language used to describe character development and to differentiate between the main and supporting characters. Holistic judgments are also made to analyze inter-textual cohesive quality referencing, and cohesion.

To our knowledge, no study has evaluated the development of children's oral narrative abilities using a rubric designed specifically to measure aspects of macrostructure and microstructure based in discourse theory. As discourse theory transcends the use of schema theory, it may be beneficial to understand the developmental trajectory of children's narratives in their school-age years. The *Monitoring Indicators of Scholarly Language* (MISL; Gillam et al., 2017) rubric is a progress-monitoring tool that has been designed to track children's development of oral narrative skill over time and is based in discourse theory. The MISL was designed to measure stories that range from simple descriptions to complex multi-episodic narratives. Both a macrostructure and microstructure subscale are included and yield a total narrative proficiency score based in discourse theory. The macrostructure subscale accounts for the use of SGEs as well as the level to which each element is causally and temporally connected in the global organization of a story. The microstructure subscale accounts for the use of literate language features necessary to establish temporal and causal connections locally in stories. It utilizes discrete measurement criteria for the use of story grammar elements as well as the causal connections between them reflecting the nature of the interrelationships between critical episodic elements. This is achieved by removing some of the subjectivity inherent in the use of holistic judgments and making them "discreet." This is more in line with the notion of macrostructure as introduced by Kintsch and van Dijk (1978) (and others), who maintain that macrostructure is not measured by documenting the presence or absence of story elements or holistic judgments of cohesion, but rather by the causal framework that exists between them.

The MISL has been shown to be a valid and reliable measure for charting progress in oral narrative growth (Gillam et al., 2017). In the first study, the MISL was used to score stories told by 109 children with language impairments (ages 5; 7–9; 9) who participated in a normative study for the Test of Narrative Language—Second Edition (TNL-2; Gillam and Pearson, 2017). The stories elicited from the Aliens subtest were used to assess psychometric adequacy measured for inter-rater reliability, internal consistency and construct validity. The Aliens subtest is an oral narrative prompt where the child is asked to tell a story from a picture. The MISL was shown to demonstrate good inter-rater reliability for the macrostructure and the microstructure subscales (ranging from 92 to 100% for each item) and acceptable levels of both internal consistency reliability (>0.70 Cronbach's alpha) and construct validity for use in measuring overall narrative proficiency (MISL total score). It has yet to be established, however, whether each subscale element is developmentally sensitive to narrator age and if so, whether that extends beyond the elementary school-age range (5–9 years). In addition, we were interested in knowing whether MISL scores across ages reflect the same developmental stages proposed by Stein and Glenn (1979) that were supported by Berman and Nir-Sagiv (2007).

Measurement of the presence of SGEs as well as their causal relationship to one another is critical if we are to gain a more thorough understanding of a child's knowledge of narrative structure. Research has explored the role of causal connectivity in written discourse and has revealed that statements in written text that include a large number of causal connections tend to be more readily recalled (Espin et al., 2007), judged as more important by the reader (Trabasso and Sperry, 1985), and retrieved from memory more quickly (O'Brien and Meyers, 1987) than statements that have a smaller number of connections. Similar findings have been described in oral discourse tasks (Cevasco and van den Broek, 2008). Though previous summaries of child's narrative development (Stein and Glenn, 1979; Berman and Nir-Sagiv, 2007) have

reported children incorporating basic episodes in their stories around ages of 7–8 years old, we predict that MISL scores would reflect a later timeline, as they require the narrator to explicitly indicate the causal relationships between SGEs, not just simply state events in a logically ordered sequence (as is the case in schema theory-based assessment tools). We therefore hypothesize that children in this age range who may have less knowledge of narrative structure, as well as less well-developed language abilities, may frequently fail to produce basic episodes with explicitly stated causal connections between initiating events, actions and consequences. To determine the typical age at which both basic and elaborate narratives were produced based on discourse theory criteria, narratives produced by children from a larger age range, including older school-age children (4–15 years), were evaluated using the MISL rubric.

The purpose of this project was to understand the nature of child's oral narrative development following discourse theory using a progress monitoring tool for children ages 4–15. To address this purpose, the following questions were posed:

1. Are measurements of macrostructure ability (as measured by the MISL rubric) sensitive to changes across age?
2. At what age do the majority of children in the sample achieve proficiency (i.e., a score of 2 or more) across each macrostructure element?

Materials and methods

Participants

A total of 687 narratives were analyzed in this study, which were elicited from participants drawn from the normative sample of 1,597 children in the TNL-2 (Gillam and Pearson, 2017). The participants ranged in age from 4; 0 to 15; 0 (M = 8; 9, SD = 2; 8). There was a roughly even split in the reported biological sex of participants, with 51.8% of narratives (n = 356) having been elicited from females, and 48.2% of narratives (n = 331) having been elicited from males. Samples were elicited from children whose reported ethnicity was white (86%), black or African American (9%), two or more ethnicities (2%), Asian or Pacific Islander (1.5%), American Indian, or Alaskan Native (0.6%), with the remaining 0.9% preferring to not respond. Close to one-third of the sample were identified as qualifying for free and reduced lunch programs (29.1%), with the remaining 70.9% either not qualifying or choosing to not report this information. Finally, narratives were elicited from children across different regions in the United States, including the Northeastern region (21.8%), the Southeastern region (16.7%), the Midwestern region (7.6%), and the Western region (53.9%).

Materials

The TNL-2 (Gillam and Pearson, 2017) is a standardized measure of narrative proficiency that assesses a child's comprehension and production of stories in three progressively independent contexts. The first context requires participants to listen to a story, answer questions about it, and retell the story (McDonald's subtest). Next, students are asked to listen to a story that is modeled using a set of sequenced pictures (Shipwreck subtest), answer questions about it, and then create a new account with a novel set of sequenced images (Late for School subtest story). The last context involves asking participants to listen to a story about a single picture (Treasure subtest), answer questions about it, and create a new account from a unique image (Aliens subtest). The prompt for the Aliens subtest is a novel scene that depicts an alien family that is landing in the park. Children are asked to generate a story based off of the picture prompt. The narratives for this project were elicited from the Aliens subtest of the TNL-2 assessment.

The MISL rubric was used to score the Aliens subtest story from each participant's TNL-2 assessment. The MISL includes a macrostructure and microstructure subscale. The scores from these scales are then combined to reflect an overall narrative proficiency score. Story elements are judged as absent (score of 0), emerging (score of 1), present/mastered (score of 2), or elaborated (score of 3). Scores on the MISL are awarded based on how the story elements (e.g., initiating event, action, consequence) are causally/temporally related rather than the number of times an element is observed in a narrative. A score of 0 is interpreted as evidence that the story does not contain the elements that make up a basic story episode. These stories may contain simple descriptions of objects or actions (e.g., There is a ship. They are eating). A score of 1 indicates that a story may have an emerging episodic structure (e.g., There is a girl. She is hiding in the bush). A score of 2 is interpreted as evidence that a story contains the necessary elements to constitute a basic story episode (e.g., The girl is hiding behind a bush and then jumped out to scare the aliens. She ran home to tell her parents about the aliens because she was scared). A score of a 3 indicates that the story is complex and elaborated (e.g., Jill and Jack were at the park. They hid behind the bush because the aliens landed. They decided to jump out from behind the bush to scare the aliens. After they scared the aliens, they ran home to tell their parents all about their day at the park. Their parents didn't believe their story, so they took them back to the park. When they got to the park, the aliens were gone). The macrostructure subsection of the MISL is designed to measure both SGEs and the temporal and causal connections that make the narrative both locally and globally coherent. There are seven SGEs measured in the MISL, including Character, Setting, Initiating Event, Internal Response, Plan, Action, and Consequence (see **Table 1**). Similar to the view of Stein and Glenn (1979), Character and Setting are scored individually, as they exist outside of the overall sequence

of the plot. The remaining elements comprise a chain of events that begins with Initiating Event and resolves with consequence. This causal chain is critical to maintaining the global cohesion of a narrative that allows the story recipient to construct and maintain a situation model, or a mental representation of the narrative, which underlies narrative comprehension (Zwaan et al., 1995; Graesser et al., 1997).

In order to determine whether a causal connection exists between statements in a story, a cause must come before its outcome (temporal priority), be in operation when the outcome occurs (operativity) and be necessary for the consequence to occur (necessity; Mackie, 1980; van den Broek, 1990; Zwaan et al., 1995). Children often produce stories in which the conditions for causality are not met. For example, in the story, "John went to the store to buy some food. He forgot his money." It is implied that John was unable to buy food because he did not bring his money. The conditions necessary for causality are not met in this case because, while there is a temporal order (went to store, forgot money) there is no "outcome" stated. Most narrative macrostructure scoring systems, that are based in schema theory, would not only the presence or absence of specific story elements with "more" being better than less (Berman, 1988; Strong, 1998; Boudreau and Hedberg, 1999; Miles and Chapman, 2002; Reilly et al., 2004).

While the presence or absence of story elements is part of the MISL scoring system, it also includes judgments about the causal nature of the events in the story. The conditions of causality in scoring story episodes is reflected in MISL scoring by utilizing an interdependent scoring system between initiating event, internal response, plan, action, and consequence. The minimal score that indicates the conditions of causality are met is a 2 for each of these items. For example, if the story stated:

John went to the store to buy groceries. He forgot his money, so he was not going to be able to buy his food. He decided to call his mother and ask her to bring him some money so he would be able to buy his groceries. He called his mother, and she was happy to bring him some money. After John's mom brought him money, he finished his grocery shopping and came home to make his mom dinner to thank her for saving the day.

The initiating event in the story was the problem of John not being able to buy food without money. He then called his mother [action causally related to buying groceries (initiating event)], requested funds (action), and received funds (action). John then bought groceries (consequence) because that is what he originally came to the store to do (initiating event). Schema systems might give credit for the presence or absence of the initiating events, action, and consequences because they are stated in the story. For example, an action might be identified if a story contained the sentence "The girl ran over to her mother." However, in order for this statement to earn a score of 2 for attempt using the MISL, it would need to be clearly tied to an initiating event such as, "The girl ran over to her mother because she was afraid of the thunder." In the previous sentence, a score of 1 would be given for the sentence, "The girl ran over to her mother" using the MISL because there is no "clear link to an initiating event" using causal language.

General procedures

Story transcription

Stories were recorded on portable digital audio recorders and transcribed verbatim by research assistants who were blind to the purpose of the study. *Systematic Analysis of Language Transcripts* conventions were used to code each utterance (*SALT*; Miller and Iglesias, 2019). The utterances were segmented into communication units (C-units) consisting of an independent main clause and phrases or clauses subordinated to it. Each transcript was reviewed by a second research assistant for spelling, mazing, morpheme segmentation and utterance segmentation. Transcription disagreements were addressed by both transcribers who listened to the digital recording together and discussed the differences until a resolution was reached. Reliability between primary and secondary transcribers was calculated on 20% of the data. The total number of C-units and mazes (i.e., false starts, revisions) were calculated, and the number of discrepancies were determined. The discrepancies were then subtracted from the total number of C-units and mazes and a percentage agreement was calculated. Reliability was 96.7% for C-unit segmentation and 96.1% for identification and coding of mazes.

Monitoring indicators of scholarly language training

Research assistants met with the first author to review the subscales, definitions and scoring criteria of the MISL using example stories. Twenty stories that represented a variety of story types and quality levels were selected for use in MISL training. Research assistants were given five stories at a time to score. After they were scored, the research assistants met with the first author to discuss the scores and the reasoning behind the scoring decisions. This process was repeated until all 20 had been scored. After the training period, the research assistants were given 10 new stories to score that were not part of the TNL-2 database. These stories were used to determine when a research assistant had reached an overall and point-by-point reliability score of 80% or higher for scoring the MISL subscales. Only then were they considered to be sufficiently trained to participate in scoring stories for the study.

Two research assistants who had met these criteria and who were blind to the purpose of the study independently used the MISL rubric to score de-identified narrative transcripts. These research assistants independently scored stories in increments of 30. After each subset of 30 stories, the research assistants met together with the first author to review scores and discuss any scoring disagreements. This was done to minimize any effect of

TABLE 1 Macrostructure subscale story elements and scoring criteria.

Story element	0 (not present)	1 (emerging)	2 (mastery)	3 (elaborated)
Character	No main character is included or an ambiguous pronoun is used to reference a person	Includes at least one main character by using a non-specific label with a determiner (e.g., the boy, a girl)	Includes at least one main character that is referenced to using a proper noun	Includes more than one main character using proper nouns
Setting	No reference to a location or time is used	Only references to a general place or time is included (this reference is not necessarily related to the story)	Reference to a specific time or place that is related to the story is included	A reference to the place are created using proper nouns, and a reference to a specific time are included
Initiating Event	No indication of an initiating event—series of descriptions	Initiating event is stated, however, this event does not motivate actions from the characters	One initiating event is stated that motivates actions from the main characters	Two or more initiating events are included that motivate separate actions from the main characters
Internal Response	No feelings from the characters are stated	Feelings from the characters are stated, however, there is not clear relationship to the initiating event.	Feelings are stated that is clearly related to the initiating event	Multiple instances of feelings are stated that are clearly related to the initiating event.
Plan	No statement is included that describes the character's plan to take action	Statements about plans to take action are included, however, these plans are not directly related to the initiating event.	One statement depicting a plan is included that is directly related to the initiating event.	Multiple statements about plans the characters have to take action are included that are directly related to the initiating event.
Attempt	No actions/attempts are taken by the characters	There is use of action verbs in descriptive sequences that do not have a clear link to an initiating event.	The use of action verbs in the story are clearly linked to the initiating event	A complicating action that impedes the actions characters take in response to the initiating event are included.
Consequence	There is no clear "ending" or resolution stated that is related to an initiating event	The outcome or resolution of the action is linked to another action, not the initiating event	One resolution of actions stated that is directly related to the initiating event	Two or more outcomes are stated that are directly related to the initiating event

coder drift, which is a phenomenon resulting from systematic and predictable variation in rater decisions over time. Any differences in scores were discussed and resolved by the research assistants under the direction of the first author. Reliability on each macro- and microstructure element was calculated on the uncorrected data for each item (point by point). The number of agreements was divided by the total number of item decisions and then multiplied by 100. Reliability between primary and secondary scorers was calculated on 100% of the data for the project. Interrater reliability for MISL total scores was 85%.

Data analysis

Pearson correlation analysis was used to first establish convergent validity between the macrostructure section of the MISL rubric and the TNL-2 Aliens subtest raw production score. This step was necessary to establish the appropriateness of utilizing the normative database collected for the TNL-2 as a normative database for MISL scores. The macrostructure total score and the TNL-2 Aliens subtest raw production scores were found to have high levels of convergent validity, based on correlation analysis, $r(686) = 0.766$, $p < 0.001$, indicating that the normative sample for the TNL-2 could adequately serve as a normative sample for the MISL macrostructure.

Research question one, which aimed to determine the sensitivity of each macrostructure element to the age of narrator was addressed through an ordinary least squares regression where age predicted total macrostructure score, followed by a series of *post hoc* ordinal logistic regressions (OLR) where age predicted each individual MISL rubric element. OLR was utilized to capture the ordinal nature of the MISL scores, which are on a scale of 0–3, with each score representative of a different level of narrative proficiency. Use of a generalized linear modeling method like OLR was necessary, as both ordinary least squares and analysis of variance assume a continuous dependent variable with normally distributed residuals. In each OLR, age of narrator predicted each individual macrostructure element score (Character, Setting, Initiating Event, Plan, Internal Response, Action, Consequence) for a total of seven models. Beta coefficients were converted to odds-ratios for ease of interpretation.

To address research question two, which was to evaluate the age at which the majority of the children in the sample had proficient scores (i.e., a score of two or higher) for each macrostructure element, descriptive statistics were utilized. Mainly an evaluation of the modal score for each age (separated by year) was evaluated for each macrostructure element.

Results

The ordinary least squares linear regression indicated that age of narrator was a significant predictor of total macrostructure score β = 0.97, $t(687) = 19.52$, $p < 0.001$, meaning that each 1-year increase in age was associated with a 0.97 point increase in macrostructure total score (see Table 2). The R-squared value estimates that 35.64% of the variance in macrostructure total score can be accounted for by age of the narrator.

Results of each OLR model indicated that age was a significant predictor of all macrostructure elements ($p < 0.001$), whereby a positive trend was seen between the age of narrator and their score on each of the seven macrostructure elements. Odds-ratios ranged between 1.13 and 1.60, indicating that for each 1-year increase in age of narrator, the odds of receiving the next highest macrostructure score increased by 1.13–1.60 times across each of the elements. The smallest effect size was seen for internal response, however, the relationship between age of narrator and MISL score was still statistically significant, [ordered odds ratio (Estimate)] = [1.13], 95% *CI* = [1.07, 1.19], *Wald* = [4.522], $p < 0.001$. The largest effect size was seen for Consequence, where each 1-year increase in age was associated with 1.6 times increase in the odds of receiving the next score level, 95% *CI* = [1.50, 1.71], *Wald* = [13.975], $p < 0.001$. Results of each OLR model are presented in Table 3.

An array of Jitter plots depicting the distribution of individual scores for each element by age is shown in Figure 1. Modal scores (i.e., the most commonly occurring score) by age for each element are discussed in the following sections and are also depicted in Table 4. Modal scores are provided in place of the mean and standard deviation, since scores are ordinal in nature and represent different stages of SGE mastery. Scores were also not normally distributed, so the mean score for each age would not accurately represent the middle of the score distribution.

Examination of the distribution of scores in Figure 1, revealed a distinct increase in scores for character for 9- and 10-year-old children; whereby younger children ages 4–9 most frequently received (mode) a score of 1 on character. Children between the ages of 10–15 most frequently received (mode) a score of 3 for character, indicating not only proficiency for this age range, but elaboration.

The modal value for setting remained at a score of 1 across all ages, however, it can be seen in Figure 1 that the distribution of scores was more widespread from ages 10 on. This means that while 1 remained the most common setting score regardless of age, older children were more likely to include Setting at the proficient or elaborated level.

For initiating event score the modal score was consistently 2, indicating proficiency, for ages seven and older. The Jitter plot in Figure 1 shows an evident cluster of scores at 3 for initiating event from age 8 and older, and a cluster of scores 0 and 1 for ages 4–7, with 2 remaining the most frequent initiating event score across all ages.

The modal score for internal response was 0 for each age apart from the 12 and 15-year-old group. As can be seen in the Jitter plot for internal response in Figure 1, the largest cluster of scores across ages was 0, however, there was a smaller number of scores at 0 from ages 12 on. This finding indicated that while it was common for narrators to exclude the use of internal response in their stories, there was greater likelihood for its inclusion at later ages.

For plan, there was a clear increase at ages 9 and 10 in its presence and sophistication in children's stories. Prior to that, for ages 4–9 the most frequent score for plan was 0. By the time students reached ages 10–15 the most frequent score for plan was 2, indicating proficiency in using the story element causally to indicate intentions of characters. The Jitter plot of plan in Figure 1 reflects these clusters, in addition to showing a small cluster of scores at 1 for the middle age range and a sparse cluster of scores at 3 in the older age range.

Following plan, action had a modal score of 2 across the majority of the age-range included in the sample (7; 0–15; 0). As can be seen in the Jitter plot, there appeared to be a greater spread in scores for narratives elicited from children between 5; 0 and 8; 0, with a roughly even spread amongst scores of 0, 1, and 2 for this age-range. There is a clearer band of scores at 2 points from ages 9; 0 to 15; 0, potentially indicating more common usage of causally connected actions at around 8–9 years of age.

Finally, the most evident break in scores could be seen for consequence, whereby there was a clear change from the absence of consequence from stories (score of 0) to the presence of consequence at the level of proficiency (score of 2) or elaboration (score of 3) at age 9 and older. As can be seen in the Jitter plot

TABLE 2 Macrostructure total score predicted by age.

	Estimate (β)	Std. Error	*t*-value	*p*-value
(Intercept)	0.574	0.456	1.26	0.208
Age	0.972	0.05	19.52	<0.001***

Statistical significance is indicated by *** = < 0.001; $R^2 = 0.355$.

TABLE 3 Results of OLR for macrostructure element scores by age.

Model	Estimate (SE)	Odds-ratio [CI]	Wald	*p*-value
Character ~ Age	0.40 (0.03)	1.49 [1.39, 1.60]	11.3	<0.001***
Setting ~ Age	0.28 (0.03)	1.33 [1.24, 1.42]	8.51	<0.001***
IE~ Age	0.41 (0.03)	1.50 [1.41, 1.60]	12.74	<0.001***
IR~ Age	0.12 (0.03)	1.13 [1.07, 1.19]	4.52	<0.001***
Plan ~ Age	0.22 (0.03)	1.25 [1.18, 1.33]	7.6	<0.001***
Action ~ Age	0.35 (0.03)	1.42 [1.33, 1.51]	11.01	<0.001***
Con ~ Age	0.47 (0.03)	1.60 [1.50, 1.71]	13.97	<0.001***

Statistical significance is indicated by *** = < 0.001. IE, Initiating Event; IR, Internal Response; Con, Consequence.

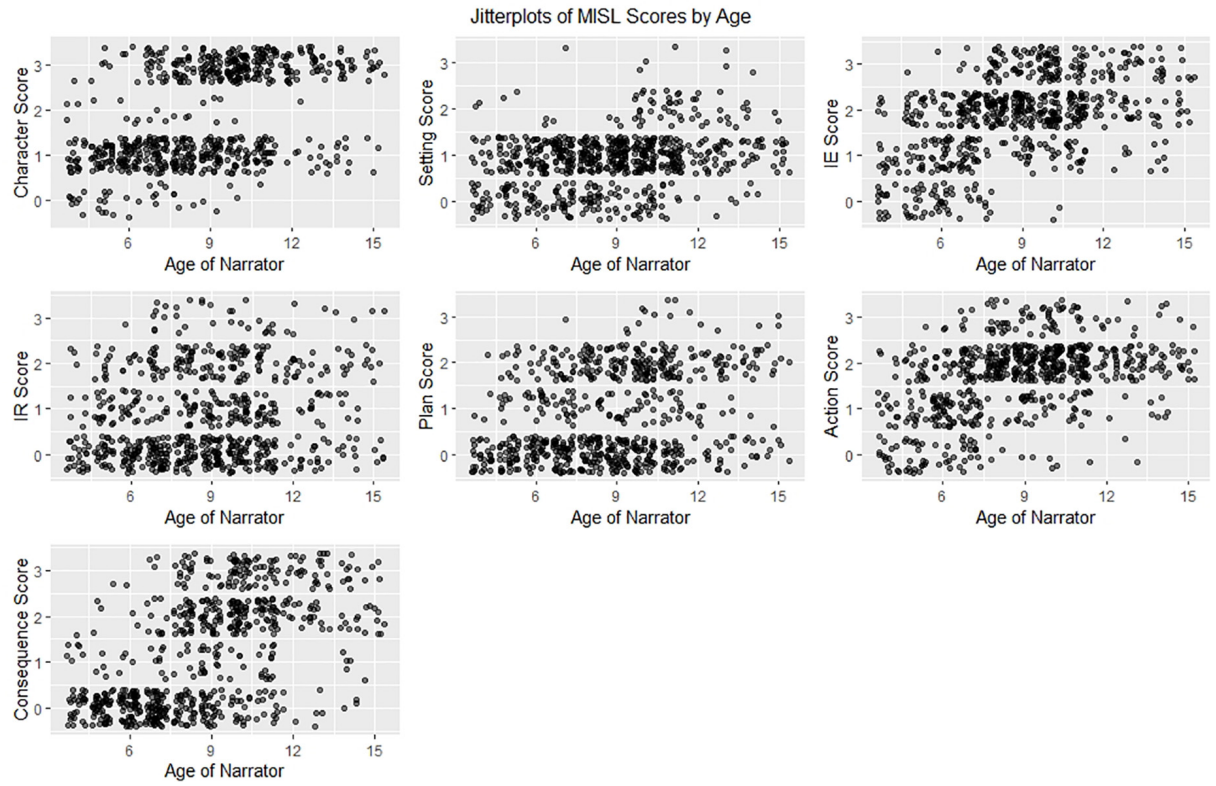

FIGURE 1
IE, initiating event; IR, internal response.

TABLE 4 Modal scores for macrostructure element by age.

	Age of narrator											
	4	5	6	7	8	9	10	11	12	13	14	15
Char	1	1	1	1	1	1	3	3	3	3	3	3
Sett	0	1	1	1	1	1	1	1	1	1	1	1
IE	0	2	1	2	2	2	2	2	2	3	3	2
IR	0	0	0	0	0	0	0	0	2	0	0	2
Plan	0	0	0	0	0	0	2	2	2	2	2	2
Act	0	0	1	2	2	2	2	2	2	2	2	2
Con	0	0	0	0	0	2	2	2	3	3	3	2

Cells highlighted in gray represent where a score ≥ 2 are consistent across increasing ages, indicating proficiency (2) or mastery (3).

for consequence (see Figure 1), the cluster of scores indicates a positive trend whereby older children received higher scores on consequence than younger children.

Discussion

The purpose of this study was to understand the nature of oral narrative macrostructure development in school-age children using a metric based in discourse theory that defines macrostructure as the story elements and the causal connections between them. This is a departure from prior work based in schema theory that has quantified oral macrostructure abilities by noting the presence or absence of story elements or making holistic judgments about the quality and developmental level of an oral narrative. The MISL incorporates discrete criteria for measuring story elements and their causal connections which removes some of the subjectivity of these earlier rubrics. Our first aim was to determine whether the MISL was sensitive to differences in oral macrostructure abilities across age using

this newer approach. Results indicated that there was strong convergent validity between the TNL-2 Aliens subtest score and the MISL score, $r(686) = 0.766$, $p < 0.001$. Results of both the simple regression on total macrostructure score and the series of ordinal regression analyses for each macrostructure element indicated that age was a significant predictor of the scores children received. Collectively, these results suggest that the MISL is a developmentally valid measure of oral narrative production abilities.

The second research question asked at what age the majority of children in the sample used each macrostructure element in oral narratives. Earlier studies used simple counts and holistic judgments of story elements to measure the quality or developmental level of narratives. The MISL requires that causal connections be explicitly stated in order for specific story elements related to the creation of a complete or complex episode (initiating event, internal response, plan, attempt, consequence) to be given a score of 2 or higher. Further, scores of 2 for character and setting are not given unless 2 or more examples of these story elements are stated in a story. This imposes a more rigid requirement on the scoring of episodes than those based in schema theory. Therefore, we expected to find that our developmental trajectory for the use of macrostructure elements might reflect a lengthier timeline than earlier studies. Following schema theory, Stein and Glenn (1979) asserted that preschool-age children reach the developmental milestone of telling stories that describe characters and list actions chronologically. These findings were supported by Berman and Nir-Sagiv (2007). Consistent with this research, the preschoolers who participated in our study were observed to include characters in their stories. For example, 4-year-olds demonstrated a modal score of 1 for character, and 0 for all other SGEs ($n = 32$).

It was not until age 6 in our sample that we observed higher modal scores of 2 for character, setting and action ($n = 64$) at which time children were describing characters by name and clearly attributing the actions they described in their stories to the characters they introduced. Our observation with regard to this finding is that schema frameworks that do not discreetly measure causality between story elements may be associated with "earlier" achievements in the use of story elements than those based in discourse theory. Similarly, Stein and Glenn (1979) asserted that by age six children typically tell stories that include the aims or intentions of their characters (plans). This was supported by John et al. (2003) where children in their sample were including internal responses and aims of the characters in story retells by the time they were 7-years-old. Using the more rigid criteria imposed by characterization of macrostructure as "including causal connections" our 6-year-olds were not observed to include clearly aligned plans and intentions in their stories, with most scoring a 0 or 1. Scores of 1 would indicate a "planning word" was used (e.g., thought, decided) but without a clear causal relationship to the character, it would not be given a score of 2 which was required to meet our definition of whether the story element was "present." The achievement of "complete episode" was reported by Stein and Glenn (1979) and Berman and Nir-Sagiv (2007) between the ages of 7–8 in which children were reported to tell stories that included an initiating event, action, and consequence. Our findings were that the three critical elements defining the achievement of complete episode (initiating event, action, consequence) occurred at 9 years of age. The MISL rubric requires that all three elements (i.e., initiating event, action, consequence) be clearly and specifically connected to each other using specific language (e.g., because, so). Following these criteria, the emergence and stabilization of consequence was most impacted. Initiating event and action stabilized slightly earlier in the current data sample. Our effect size estimates indicated that for each 1-year increase in age, the score for the use of a specific element increased. Thus, as children aged, they were increasingly better at using language to link story elements together using causal language. For example, younger children (i.e., 4 and 5-year-olds) produced stories like, "The car was crashing. The people were walking by the car." Stories like this received scores of 1 for initiating event and action because a possible initiating event (e.g., The car was crashing) was stated, and an action was stated (e.g., The people were walking by the car.), however, the events were not causally and temporally related. Whereas children who were 9 years of age produced stories like:

The car was about to crash into the big hole, so the people inside started to scream. Then, they pressed on the brakes and turned the wheel to get away from the hole. They missed the hole, and everyone was safe.

Stories like this received a score of 2 for initiating event, action, and consequence—indicating that the three critical elements of a story were included, and they were causally related. In contrast, using a rubric based in schema theory, younger children would receive the same scores as children who were 9 years old in our sample because they would receive scores that reflected whether a story grammar element was present or not.

It was not until age 10 that we observed scores of 2 for the story elements of internal response and plan ($n = 273$). Stories included words that might be associated with planning or feelings the characters may have had, however, students were not observed to consistently use causal language to connect them to the basic episode until much later than reported in earlier studies (Stein and Glenn, 1979; John et al., 2003; Berman and Nir-Sagiv, 2007). It is well supported that individuals from a very young age regularly pay attention to goal motivated actions, plans, and internal responses in the stories they hear or read and tend to include those elements in the stories they create on their own (Lynch and van den Broek, 2007). However, children in the current sample were not shown to consistently use literate language features to establish causal connections between an initiating event and a plan until the age of 10. A child in our sample who was 10 years of age might produce a story like:

The car was about to crash into the big hole. The people inside of the car were scared. Then, they decided to press on the brakes and turned the wheel to try and get away from the hole and that is what they did. They missed the hole, and everyone was safe.

This story would have received a score of 2 for plan because it was temporally related to the initiating event. A score of 1 would have been rewarded for internal response because a feeling was stated. The scores using the MISL for children this age reflect the emergence/presence of some of the story grammar elements (e.g., internal response), and mastery of others (e.g., initiating event, action, plan, consequence). In contrast, using a rubric based in schema theory, the children would have received scores that reflected mastery of all of the story grammar elements by this age because the child included at least one example of the element in their story.

Finally, the emergence of complex episodes in oral narratives was reported by Stein and Glenn (1979) asserted that by the time a child is 11 years of age produce stories that are intricate and include an elaboration of the complete episode. This also has been supported by a variety of studies looking at the complexity of narratives in written contexts (Dockrell and Connelly, 2016; Jagaiah et al., 2020). Remember that elaboration occurs when a child includes multiple episodes with and more than one plan, action sequence. This was supported in oral narratives by the work of John et al. (2003), where 11-year-old children were found to include elements like internal response at a higher rate than their younger peers. In our sample, it was only children 13 years of age and older ($n = 67$) that were shown to consistently elaborate on their story and include complex episodes in oral narratives. A child who was 13 in our sample might produce a story like:

The white Jeep was about to crash into the big hole in the desert. The people inside of the car were scared. Then, the driver John said, "Hey. Everyone stop screaming so I can think." He decided to press on the brakes and turned the wheel to try and get away from the hole and that is what they did. They missed the hole, and everyone was safe. Then, all of a sudden, a huge thunderstorm came, and rain started falling fast. Everyone was getting wet, so they decided to drive and find a rock to hide under. That's what they did. They found a rock and waited until the thunderstorm ended to go home.

This story would have received scores of a 3 for initiating event, plan, action, and consequence because the child included more than one complete story episode where these elements were causally and temporally related. In contrast, using a rubric based in schema theory, a child that produced a more complex story would have received scores similar to those observed at ages 9 or 10 because they would have only received a point based off of the presence/absence of the story grammar elements.

It was not until the age of 15 that we observed scores that reflected "mastery" for the use of internal response (feelings). A child that was 15 years of age may have produced a story like:

The white Jeep was about to crash into the big hole in the desert. The people inside of the car were scared so they started to scream and panic. Then, the driver John said, "Hey. Everyone stop screaming so I can think." He decided to press on the brakes and turned the wheel to try and get away from the hole and that is what they did. They missed the hole, and everyone was safe. John felt relieved.

A child who produced this story would have received a score of 2 for internal response because the relationship to the feelings was explicitly stated and related to the initiating event. Prior research that has reported the earlier use of internal response at the age of 9 was conducted using story retell data (Berman and Slobin, 1994). Research has demonstrated that having an adult model in a story retell task has benefited the narrative performance of typically developing children for sentence complexity and story macrostructure (Sheng et al., 2020). In addition, research has demonstrated that both monolingual and bilingual children include more content in their stories when retelling a story vs. telling a unique story from a picture (Schneider and Dube, 2005; Lucero and Uchikoshi, 2019). The current research utilized story tells which may require more sophisticated language ability. This may have contributed to the findings that the mastery of this element was not found until the children were 15 years old. It could be that the nature of the task (i.e., creating a story) made it more difficult for the children to utilize complex language to create temporal and causal connections in their stories related to the use of internal response.

Clinical implications

A child's ability to successfully produce a narrative is an important developmental milestone for school-age children (Hughes et al., 1997). As narratives are complex in nature, they can be used as a measure of language ability throughout development (Hudson and Shapiro, 1991; Hughes et al., 1997; National Governors Association Center for Best Practices and Council of Chief State School Officers, 2010; Ukrainetz, 2015; Petersen et al., 2020). Many studies have discussed the usefulness of a schema theory-based approaches which employ measurement of SGEs to examine narrative ability (Stein and Glenn, 1979; Merritt and Liles, 1987; Berman and Nir-Sagiv, 2007; Bitetti and Hammer, 2021). Not surprisingly, many narrative interventions have been designed that focus on the explicit teaching of SGEs to students who are delayed in their narrative language abilities (for reviews with examples see Petersen, 2011; Favot et al., 2021; Pico et al., 2021). However, measuring narrative discourse abilities using rubrics based in schema theory may not provide clinicians with a complete picture of the underlying language abilities a child has to bring to the "narrative production table." This has the potential to result in the use of narrative interventions that do not address the

nature of the difficulties children may experience in becoming proficient in narrative comprehension and production (oral and written). Studies of written discourse have consistently shown that statements in stories that have a large number of causal connections tend to be judged more important, recalled more frequently and retrieved more quickly, than stories with fewer causal connections (Trabasso and Sperry, 1985; O'Brien and Meyers, 1987; Espin et al., 2007). These findings have also been reported for oral discourse (Cevasco and van den Broek, 2008). Current rubrics and narrative macrostructure scoring systems that focus almost entirely on the presence or absence of story elements while asking the rater to make a holistic judgment about whether the story is also "cohesive" in nature may not capture this important aspect of narrative ability.

Limitations and future directions

There were several limitations of our study that are important to consider in the interpretation of the findings. One limitation is the differences in sample sizes at different ages. Participants in this study were drawn from the normative sample for the TNL-2, meaning that we had a larger number of children toward the middle of the age distribution than on the edges of the distribution (i.e., the youngest and oldest ages in the sample). However, because age was normally distributed, we found it appropriate to conduct an ordinal logistic regression to analyze our data and account for differences at each score level by age. Additionally, a potential limitation is found in the generalizability of our findings. The original data used in our analyses came from participants in a few different locations in the United States. Given that the majority of the participants were Caucasian, it is difficult to determine if our results would generalize to children of other ethnic and cultural backgrounds. The benefit of narrative sample analysis, however, is that their use tends to be more sensitive to such differences in backgrounds of participants than standardized assessments (MacLachlan and Chapman, 1988; Dollaghan et al., 1990; Leadholm and Miller, 1992; Wagner et al., 2000; Nippold et al., 2014). Still, additional analyses on a more diverse population of children are needed to better generalize these results to the population of school-age children. As culturally and ethnically diverse populations grow in the United States, it would be beneficial to understand whether results for narrative production would vary across diverse backgrounds.

Finally, stories for this study were from the Aliens subtest from the TNL-2. This study has evidence for the use of this rubric for a spontaneous story-generation prompt. It may be necessary to conduct studies to understand the use of the MISL on stories produced from different elicitation contexts to continue to explore its validity. In the future, it would be beneficial to explore the validity of the MISL across different narrative elicitation contexts. Results might differ for contexts such as story-retell or personal narratives, which are also commonly used in assessment. In addition, it may be important to understand the use of the MISL with children of differing language abilities, including those who are at-risk for language impairment. This would increase our ability to understand differences in narrative production abilities of typically developing children and those who are at-risk. That knowledge may lead to stronger evidence for the use of interventions targeted at increasing narrative production abilities of children who are at-risk for language impairment.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Internal Review Board at Utah State University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

MI-A contributed to the conception and design of this study and wrote the first draft of this manuscript. CF was main data analyst on project and wrote sections of the manuscript. SH contributed to edits of the first draft of the manuscript. SG was co-PI on data collected for this project and contributed to editing. RG was co-PI on data collected for this project. All authors will contribute to revisions of the submitted version of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Applebee, A. N. (1978). *The Child's Concept of a Story: Ages two to Seventeen*. Chicago: University of Chicago Press.

Berman, R. A. (1988). On the ability to relate events in narrative. *Discourse Process.* 11, 469–497. doi: 10.1080/01638538809544714

Berman, R. A., and Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Process.* 43, 70–120. doi: 10.1080/01638530709336894

Berman, R. A., and Slobin, D. I. (1994). "Narrative Structure," in *Relating Events in Narrative: A Crosslinguistic Developmental Study*, eds R. A. Berman and D. I. Slobin (Hillsdale, NJ: Lawrence Erlbaumn Associates, Inc), 39–84.

Bitetti, D., and Hammer, C. S. (2021). English narrative macrostructure development of Spanish–English bilingual children from preschool to first grade. *Am. J. Speech Lang. Pathol.* 30, 1100–1115. doi: 10.1044/2021_AJSLP-20-00046

Botting, N. (2002). Narrative as a clinical tool for assessment of linguistic and pragmatic impairments. *Child Lang.Teach. Therapy* 18, 1–22. doi: 10.1191/0265659002ct224oa

Boudreau, D. M., and Hedberg, N. L. (1999). A comparison of early literacy skills in children with specific language impairment and their typically developing peers. *Am. J. Speech Lang. Pathol.* 8, 249–260. doi: 10.1044/1058-0360.0803.249

Carvalhais, L., Limpo, T., and Pereira, L. A. (2021). The contribution of word-, sentence-, and discourse-level abilities on writing performance: A 3-year longitudinal study. *Front. Psychol.* 12:668139. doi: 10.3389/fpsyg.2021.668139

Catts, H. W., Fey, M. E., Tomblin, J. B., and Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *J. Speech Lang. Hear. Res.* 45, 1142–1157. doi: 10.1044/1092-4388(2002/093)

Cevasco, J., and van den Broek, P. (2008). The importance of causal connections in the comprehension of spontaneous spoken discourse. *Psicothema* 20, 801–806.

Crais, E. R., and Lorch, N. (1994). Oral narratives in school-age children. *Top. Lang. Disord.* 14, 13–28. doi: 10.1097/00011363-199405000-00004

Dockrell, J. E., and Connelly, V. (2016). "The relationships between oral and written sentence generation in English speaking chidlre: The role of language and literacy skills," in *Written and Spoken Langauge Development Across the Lifespan*, eds J. Perera, M. Aparici, E. Rosado, and N. Salas (Cham: Springer), 161–177. doi: 10.1007/978-3-319-21136-7_11

Dockrell, J. E., Connelly, V., Walter, K., and Critten, S. (2014). Assessing children's writing products: The role of curriculum-based measures. *Br. Educ. Res. J.* 41, 575–595. doi: 10.1002/berj.3162

Dollaghan, C. A., Campbell, T. F., and Tomlin, R. (1990). Video narration as a language sampling context. *J. Speech Hear. Disord.* 55, 582–590. doi: 10.1044/jshd.5503.582

Espin, C. A., Cevasco, J., van den Broek, P., Baker, S., and Gersten, R. (2007). History as narrative: The nature and quality of historical understanding for students with LD. *J. Learn. Disabil.* 40, 174–182. doi: 10.1177/00222194070400020801

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

Gillam, R. B., and Pearson, N. A. (2017). *TNL-2: Test of Narrative Language – Second Edition*. Austin: TX: Pro-ed.

Gillam, S. G., Gillam, R. B., Fargo, J. D., Olszewski, A., and Segura, H. (2017). Monitoring indicators of scholarly language: A progress-monitoring instrument for measuring narrative discourse skills. *Commun. Disord. Quarterly* 38, 96–106. doi: 10.1177/1525740116651442

Graesser, A. C., Millis, K. K., and Zwaan, R. (1997). Discourse comprehension. *Ann. Rev. Psychol.* 48, 163–189. doi: 10.1146/annurev.psych.48.1.163

Graham, S., Gillespie, A., and McKeown, D. (2013). Writing: Importance, development, and instruction. *Read. Writ.* 26, 1–15. doi: 10.1007/s11145-012-9395-2

Greenhalgh, K. S., and Strong, C. J. (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Lang. Speech Hear. Ser. Sch.* 32, 114–125. doi: 10.1044/0161-1461(2001/010)

Hedberg, N. L., and Westby, C. E. (1993). *Analyzing Storytelling Skills: Theory to Practice*. Tucson, AZ: Communication Skill Builders.

Heilmann, J., Miller, J. F., Nockerts, A., and Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *Am. J. Speech Lang. Pathol.* 19, 154–166. doi: 10.1044/1058-0360(2009/08-0024)

Hudson, J. A., and Shapiro, L. R. (1991). "From knowing to telling: The development of children's scripts, stories, and personal narratives," in *Developing Narrative Structure*, eds A. McCabe and C. Peterson (Mahwah, NY: Lawrence Erlbaum Associates, Inc), 89–136.

Hughes, D., McGillivray, L., and Schimidek, M. (1997). *Guide to Narrative Language: Procedures for Assessment*. Eau Claire: Thinking Publications.

Jagaiah, T., Olinghouse, N. G., and Kearns, D. M. (2020). Syntactic complexity measures: Variation by genre, grade-level, students' writing abilities, and writing quality. *Read. Writ.* 33, 2577–2638. doi: 10.1007/s11145-020-10057-x

John, S. F., Lui, M., and Tannock, R. (2003). Children's story retelling and comprehension using a new narrative resource. *Can. J. Sch. Psychol.* 18, 91–113. doi: 10.1177/082957350301800105

Justice, J. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., and Gillam, R. B. (2006). The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative performances. *Am. J. Speech Lang. Pathol.* 15, 177–191. doi: 10.1044/1058-0360(2006/017)

Kintsch, W., and van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychol. Rev.* 85, 363–394. doi: 10.1037/0033-295X.85.5.363

Leadholm, B. J., and Miller, J. F. (1992). *Language Sample Analysis: The Wisconsin Guide*. Wisconsin: Wisconsin Department of Public Instruction.

Liles, B. (1993). Narrative discourse in children with language disorders and children with normal language. *J. Speech Lang. Hear. Res.* 36, 868–882. doi: 10.1044/jshr.3605.868

Liles, B., Duffy, R. J., Merritt, D. D., and Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *J. Speech Lang. Hear. Res.* 38, 415–425. doi: 10.1044/jshr.3802.415

Lucero, A., and Uchikoshi, Y. (2019). Narrative assessments with first grade Spanih-English emergent bilinguals: Spontaneous versus retell conditions. *Narrat. Inq.* 29, 137–156. doi: 10.1075/ni.18015.luc

Lynch, J. S., and van den Broek, P. (2007). Understanding the glue of narrative structure: Children's on- and off-line inferences about character's goals. *Cogn. Dev.* 22, 323–340. doi: 10.1016/j.cogdev.2007.02.002

Mackie, J. L. (1980). *The Cement of the Universe: A Study of Causation*. Oxford: Oxford (Clarendon) University Press. doi: 10.1093/0198246420.001.0001

MacLachlan, B. G., and Chapman, R. S. (1988). Communication breakdowns in normal and language learning-disabled children's conversation and narration. *J. Speech Hear. Disord.* 53, 2–7. doi: 10.1044/jshd.5301.02

Mandler, J. M., and Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cogn. Psychol.* 9, 111–151. doi: 10.1016/0010-0285(77)90006-8

Mayer, M. (1967). *A Boy, a Dog and a Frog*. New York, NY: Dial Press.

Mayer, M. (1969). *Frog, Where Are You?*. New York, NY: Dial Press.

Mayer, M. (1974). *Frog Goes to Dinner*. New York, NY: Dial Press Books for Young Readers.

Mayer, M. (1975). *One Frog too Many*. New York, NY: Dial Press Books for Young Readers.

Merritt, D. D., and Liles, B. Z. (1987). Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension. *J. Speech Lang. Hear. Res.* 30, 539–552. doi: 10.1044/jshr.3004.539

Meyer, B. J. F. (1975). *The Organization of Prose and its Effects on Memory*. Amsterdam: North-Holland Publishing Company.

Miles, S., and Chapman, R. S. (2002). Narrative content as described by individuals with Down syndrome and typically developing children. *J. Speech Lang. Hear. Res.* 45, 175–189. doi: 10.1044/1092-4388(2002/013)

Miller, J., Andriacchi, K., DiVall-Rayan, J., and Lien, P. (2003). *Narrative Scoring Scheme*. Madison, MDN: SALT Software, LLC.

Miller, J., and Iglesias, A. (2019). *Systematic Analysis of Language Transcripts (SALT), Research Version 20 [Computer Software]*. Madison, MDN: SALT Software, LLC.

Munoz, M. L., Gillam, R. B., Pena, E. D., and Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. *Lang. Speech Hear. Ser. Sch.* 34, 332–342. doi: 10.1044/0161-1461(2003/027)

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common Core State Standards. for English Language Arts*. Washington, DC: National Governors Association Center for Best Practices.

Nicolosi, L., Harryman, E., and Kresheck, J. (2004). *Terminology of communication disorders: Speech-language-hearing*, 5th Edn. Baltimore, MD: Lippincott Williams & Wilkins.

Nippold, M. A., Frantz-Kaspar, M. W., Cramond, P. M., Kirk, C., Hayward-Mayhew, C., and MacKinnon, M. (2014). Conversational and narrative speaking in adolescents: Examining the use of complex syntax. *J. Speech Lang. Hear. Res.* 57, 876–886. doi: 10.1044/1092-4388(2013/13-0097)

O'Brien, E. J., and Meyers, J. L. (1987). The role of causal connections in the retrieval of text. *Memory Cogn.* 15, 419–427. doi: 10.3758/BF03197731

Petersen, D. B. (2011). A systematic review of narrative-based language intervention with children who have language impairment. *Commun. Disord. Quarterly* 32, 207–220. doi: 10.1177/152574010935 3937

Petersen, D. B., Spencer, T. D., Konishi, A., Sellars, T. P., Foster, M. E., and Robertson, D. (2020). Using parallel, narrative-based measures to examine the relationship between listening and reading comprehension: A pilot study. *Lang. Speech Hear. Ser. Sch.* 51, 1097–1111. doi: 10.1044/2020_LSHSS-19-00036

Peterson, C., and McCabe, A. (1983). *Developmental Psycholinguistics: Three Ways of Looking at a Child's Narrative*. New York, NY: Plenum Press. doi: 10.1007/978-1-4757-0608-6

Pico, D. L., Hessling, Prahl A, Biel, C. H., Peterson, A. K., Biel, E. J., Woods, C., et al. (2021). Interventions designed to improve narrative language in school-age children: A systematic review with meta-analyses. *Lang. Speech Hear. Ser. Sch.* 52, 1109–1126. doi: 10.1044/2021_LSHSS-20-00160

Purcell, S. L., and Liles, B. Z. (1992). Cohesion repairs in the narratives of normal-language and language-disordered school-age children. *J. Speech Lang. Hear. Res.* 35, 354–362. doi: 10.1044/jshr.3502.354

Reilly, J., Losh, M., Bellugi, U., and Wulfeck, B. (2004). Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain Lang.* 88, 229–247. doi: 10.1016/S0093-934X(03)00101-9

Roth, F. P., and Speckman, N. J. (1986). Narrative discourse: Spontaneously generated stories of learning-disabled and normally achieving students. *J. Speech Lang.-Hear. Pathol.* 51, 8–23. doi: 10.1044/jshd.5101.08

Roth, F. P., Speech, D. L., and Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *J. Educ. Res.* 95, 259–272. doi: 10.1080/00220670209596600

Rumelhart, D. E. (1975). "The active structural network," in *Explorations in cognition*, eds D. A. Norman and D. E. Rumelhart (San Francisco, CA: W. H. Freeman).

Rumelhart, D. E. (1977). *Introduction to Human Information Processing*. New York, NY: Wiley.

Schneider, P., and Dube, R. V. (2005). Story presentation effects on chidlren's retell content. *Am. J. Speech Lang. Pathol.* 14, 52–60. doi: 10.1044/1058-0360(2005/007)

Sheng, L., Shi, G., Wang, D., Hao, Y., and Zheng, L. (2020). Narrative production in Mandarin-speaking children: Effects of language ability and elicitation method. *J. Speech Lang. Hear. Res.* 63, 774–792. doi: 10.1044/2019_JSLHR-19-00087

Stadler, M. A., and Ward, G. C. (2005). Supporting the narrative development of young children. *Early Child. Educ. J.* 33, 73–80. doi: 10.1007/s10643-005-0024-4

Stein, M. L., and Glenn, C. (1979). "Analysis of story comprehension in elementary school children," in *New Directions in Discourse Processing*, Vol. 2, ed. R. O. Freedle (New York, NY: Ablex), 53–120.

Stein, N. L. (1988). "The development of children's storytelling skill," in *Child Language: A Reader*, eds M. B. Franklin and S. Barten (Oxford: Oxford University Press), 282–297.

Strong, E. J. (1998). *The Strong Narrative Assessment Procedure*. Texas: Thinking Publications.

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cogn. Psychol.* 9, 77–110. doi: 10.1016/0010-0285(77)90005-6

Trabasso, T., and Sperry, L. L. (1985). Causal relatedness and importance of story events. *J. Memory Lang.* 24, 595–611. doi: 10.1016/0749-596X(85)90048-8

Ukrainetz, T. A. (2015). "Telling a good story: Teaching the structure of narrative," in *School-Age Language Intervention: Evidence-Based Practices*, ed. T. A. Ukrainetz (Austin: Pro-ed), 335–378.

van den Broek, P. (1990). "Causal inferences and the comprehension of narrative text," in *Inferences and text Comprehension*, eds A. C. Graesser and G. H. Bower (Cambridge, MA: Academic Press), 175–196. doi: 10.1016/S0079-7421(08)60255-8

van Dijk, T. A., and Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Cambridge, MA: Academic.

Wagner, C. R., Nettelbladt, U., Sahlén, B., and Nilholm, C. (2000). Conversation versus narration in pre-school children with language impairment. *Int. J. Lang. Commun. Disord.* 35, 83–93. doi: 10.1080/136828200247269

Westerveld, M. F., Gillon, G. T., and Miller, J. F. (2004). Spoken language samples of New Zealand children in conversation and narration. *Adv. Speech Lang. Pathol.* 6, 195–208. doi: 10.1080/14417040400010140

Zwaan, R. A., Langston, M. C., and Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychol. Sci.* 6, 292–297. doi: 10.1111/j.1467-9280.1995.tb00513.x