Research in Developmental Disabilities 55 (2016) 301–311

Contents lists available at ScienceDirect

Research in Developmental Disabilities

Why do children with language impairment have difficulties with narrative macrostructure?

Elma Blom*, Tessel Boerma

Utrecht University, Department of Special Education: Cognitive and motor disabilities, Heidelberglaan 1, 3584CS Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 15 January 2016
Received in revised form 2 May 2016
Accepted 4 May 2016
Number of reviews completed is 2

Keywords:
Language impairment
Narratives
Story comprehension
Story generation

ABSTRACT

Background: Research has produced conflicting findings about the effects of language impairment (LI) on narrative macrostructure outcomes.
Aims: The present study investigated if children with LI perform weaker than typically developing (TD) controls on narrative macrostructure in different tasks, whether this changes over time and if between-group differences stem from linguistic or cognitive factors.
Methods and procedures: A group of monolingual Dutch children with LI (n = 84) and a TD control group (n = 45) were tested with a story comprehension and a story generation task. All children were five or six at wave 1 and six or seven at wave 2. Information was collected on vocabulary, grammar, verbal memory and sustained attention.
Outcomes and results: At wave 1, the LI group performed weaker than the TD group in both tasks. At wave 2, the groups performed similarly on story comprehension. On story generation, the TD group still outperformed the LI group. Sustained attention mediated the relationship between group and story generation.
Conclusions and implications: Effects of LI on narrative macrostructure are moderated by age and task and may stem from sustained attention weaknesses. These findings have implications for using narrative tasks in educational and diagnostic settings and may direct future interventions.

What this paper adds?

This paper provides, first, insight into why LI may not always impact narrative macrostructure and, second, pinpoints possible causes of the effect of LI on narrative macrostructure. It is shown that the gap between TD and LI becomes smaller as the children grow older. These findings suggest that the different findings across studies may be an effect of age and type of narrative task. This paper adds furthermore to our understanding of the mechanisms that underlie the weak narrative macrostructure performance of children with LI by performing a mediation analysis on a sufficiently large sample. This paper shows that sustained attention contributes significantly to the lower performance of children with LI in a story generation task.

* Corresponding author.
E-mail addresses: w.b.t.blom@uu.nl (E. Blom), t.d.boerma@uu.nl (T. Boerma).

http://dx.doi.org/10.1016/j.ridd.2016.05.001
0891-4222/

1. Introduction

Language Impairment (LI) is a frequently occurring disorder that affects between 5 and 7% of the population (Bishop, 2010; Tomblin et al., 1997). Children with LI have difficulties learning language without a clearly discernible cause (Leonard, 2014). Much research on LI has focused on grammar, which tends to be severely affected by the impairment, but less is known about the narrative skills of children with LI. Narrative skills comprise the ability to sequence events, understand why one event may cause another, and structure events and descriptions in such a way that a listener can comprehend the story. A good narrator uses precise vocabulary, is able to create cohesiveness through the use of linguistic markers and conveys ideas without extra-linguistic support (Paul & Smith, 1993). It is therefore not surprising that children with LI score low on narrative measures (cf. Epstein & Philips, 2009; Botting, 2002).

Yet, there are also studies that do not observe an effect of LI, in particular on narrative macrostructure measures (Iluz-Cohen & Walters, 2012; Liles, Duffy, Merritt, & Purcell, 1995; Merritt & Liles, 1987; Norbury & Bishop, 2003). Narrative macrostructure refers to the global organization of a narrative beyond the utterance level. This structure can be analyzed using a story grammar, which identifies components such as the setting and a logical event structure (e.g., initiating event, internal response, plan, action, consequence, reaction) (Stein & Glenn, 1979). Narrative macrostructure abilities allow children to generate coherent and age-appropriate extended discourse (Heilmann, Miller, Nockerts, & Dunaway, 2010), influence their academic achievements (Bishop & Adams, 1992; Bishop & Edmundson, 1987; Boudreau, 2008; Boudreau & Hedberg, 1999; Gillam & Johnston, 1992; O'Neill, Pearce, & Pick, 2004; Tabors, Snow, & Dickinson, 2001), and are a key component of social communication (Norbury, Gemmel, & Paul, 2014). They continue to develop until at least age 9 or 10 (Norbury & Bishop, 2003), and require world knowledge, genre-specific knowledge, linguistic knowledge as well as sufficient cognitive resources to coordinate and integrate these different kinds of knowledge in real-time (Colozzo, Gillam, Wood, Schnell, & Johnston, 2011; Liles, 1993; Owens, 1996).

The first aim of the study was to investigate age effects by comparing narrative macrostructure performance of children with TD and LI across two time points. The children in this study were tested at the beginning of elementary school when narrative skills are still developing. Both narrative comprehension and generation were tested. Investigating two time points and comparing comprehension and generation are relevant for determining if effects of LI are dependent on age and type of narrative task. The second goal of the study was to determine if effects of LI on narrative macrostructure are caused by linguistic impairment, cognitive delays or both. This issue is relevant in light of the different abilities that are needed for performing well on tasks testing narrative macrostructure and findings showing that children with LI do not only have linguistic but also cognitive weaknesses (Ebert & Kohnert, 2011; Montgomery, Magimairaj, & Finney, 2010).

1.1. Narrative abilities of children with LI

Several studies have found that children with LI perform less well than children with TD on tasks testing narrative macrostructure (Bishop & Edmundson, 1987; Bishop & Donlan, 2005; Botting, 2002; Cleave, Girolametto, Chen, & Johnson, 2010; Epstein & Philips, 2009; Soodla & Kikas, 2010; Wetherell, Botting, & Conti-Ramsden, 2007). They use fewer complete episodes that consist of a logical sequence of events (such as goal-attempt-outcome) (Bishop & Donlan, 2005; Gillam & Carlile, 1997; Merritt & Liles, 1987; Miranda, McCabe, & Bliss, 1998; Wright & Newhoff, 2001) and evaluative statements that refer to the internal states of the protagonists (Bishop & Donlan, 2005; Manhardt & Rescorla, 2002; Reilly, Losh, Bellugi, & Wulfeck, 2004). Studies that investigated story comprehension observe that children with LI have difficulties with questions that assess the understanding of causal relationships between events in the narrative (Merritt & Liles, 1987) and inferencing questions, but not with factual or literal questions (Bishop & Adams, 1992; Dodwell & Bavin, 2008; Merritt & Liles, 1987).

Not all studies on narrative macrostructure observe a negative effect of LI (Iluz-Cohen & Walters, 2012; Liles et al., 1995; Merritt & Liles, 1987; Norbury & Bishop, 2003). The reasons for this are manifold. Well-structured narratives require (non-linguistic) skills that are not always impaired in children with LI (Liles et al., 1995; Norbury & Bishop, 2003). Furthermore, studies have used different inclusion criteria for LI, coding systems to analyze the plot structure may not always be sufficiently fine-grained to detect effects of LI (Duinmeijer, de Jong, & Scheper, 2012) and some components of narrative macrostructure (e.g., the ability to provide an adequate story ending) develop late in children with TD resulting in overall floor performance (Norbury & Bishop, 2003). The type of narrative task may also influence the effect of LI (Andreu, Sanz-Torrent, Guardia Olmos, & MacWhinney, 2011; Duinmeijer et al., 2012; Merritt & Liles, 1989). For instance, Merritt and Liles (1989) conclude that story retelling may provide more possibilities to detect effects of LI than story generation because it elicits more story components. Another possibility, which is relatively underresearched, is that the effects of LI emerge or fade as an effect of development.

The first aim of the study was to investigate age effects by comparing narrative macrostructure performance of children with TD and LI across two time points. A model story was used to assess narrative comprehension and generation. The model story presented children with an example without burdening their verbal short-term memory. In so doing, it could be a more reliable test of children's ability to generate a story than a retelling task (Duinmeijer et al., 2012) and may elicit more story components than a generation task without an example (Merritt & Liles, 1989). The study builds on recent research by Boerma, Leseman, Timmermeister, Wijnen and Blom (2016). For the purpose of the present study, data from more children were analyzed and data from the same children were analyzed at two time points.

1.2. Language, cognition and narrative macrostructure: implications for LI

The second aim of the study was to shed light on why children with LI may experience difficulties with narrative macrostructure. Linguistic factors may play a role, as some studies have found associations between narrative macrostructure outcomes and language level (Heilmann, Miller, & Nockerts, 2010; Heilmann, Miller, Nockerts, & Dunaway, 2010). Most typical of the impairment are grammar weaknesses (Clahsen, 1989; Rice & Wexler, 1996). In the domain of syntax, these weaknesses are reflected in difficulties with complex clauses (Friedmann & Novogrodsky, 2004; Leonard, 1995). Complex clauses are relevant for narrative macrostructure because they are often used to describe mental states and relations between events (De Villiers, 2000). Limited vocabulary knowledge (Rice & Hoffman, 2015) could furthermore reduce the level of detail of LI children's stories and hinder their story comprehension. However, some studies observe only modest relationships between language level and narrative macrostructure (Norbury & Bishop, 2003; Uccelli & Paíez, 2007). Moreover, research with bilingual children has shown that narrative macrostructure is relatively insensitive to limited language proficiency (Boerma et al., 2016; Hipfner-Boucher et al., 2014; Pearson, 2002), which may suggest that linguistic factors will not necessarily prevent children with LI from comprehending and generating structurally complex narratives.

Besides language limitations, children with LI often suffer from reduced sustained attention (*see* Ebert & Kohnert, 2011; for a meta-study) and verbal short-term and working memory (*see* Montgomery et al., 2010; for an overview). For the purpose of the present study it is relevant that both sustained attention and verbal memory are implicated in children's performance on narrative tasks. Keeping focused throughout a narrative is needed for ordering the events in a logical manner and understanding relations between events. Using a story generation task, Duinmeijer et al. (2012) indeed found significant correlations between sustained attention and narrative macrostructure provided that the narrative task was sufficiently long. Verbal memory limitations are linked to narrative macrostructure outcomes as well, but may specifically be a hindrance in story recall or retell tasks (Dodwell & Bavin, 2008; Duinmeijer et al., 2012).

In sum, both linguistic and cognitive limitations could explain why children with LI perform weaker on narrative macrostructure than TD children. While previous studies investigated correlations between linguistic and cognitive measures, on the one hand, and narrative macrostructure, on the other hand, the present study takes the next step by performing a mediation analysis in order to pinpoint which factors explain the effect of LI on narrative macrostructure.

1.3. Predictions for the present study

The first aim of this study was to compare story generation and story comprehension across TD and LI based on data collected at two waves. The study builds on a previous study by Boerma et al. (2016) who found that although the children with LI performed accurately on story comprehension, their performance was still lower than that of their TD peers who showed ceiling performance. Assuming that children with LI continue to develop, we expected that one year later the LI group may not differ from their TD peers on comprehension. For story generation, the TD children in the study by Boerma et al. (2016) did not reach ceiling at wave 1, leading to three possible outcomes for wave 2: the gap between TD and LI may be widening, remain similar or may become narrower.

The second goal of the study was to determine if effects of LI on narrative macrostructure are caused by linguistic impairment, cognitive delays or both. In order to investigate the influence of language level, several measures were included that tapped into vocabulary and grammar abilities in Dutch, which was the language of testing. The stories used in the present study consisted of three episodes and were sufficiently long to require sustained attention. Effects of verbal short-term and working memory were expected to be limited because the children did not have to retell a story. The sample in this study was relatively large and suited for a mediation analysis. Mediation analyses aim at detecting causal relationships between independent variable (Y), mediator variables (M) and dependent variable (X). One criterion for cause-and-effect (Y causes M, M causes X) is temporal precedence of the cause. The two-wave design enabled investigating relationships between linguistic and cognitive outcomes at wave 1 and narrative macrostructure outcomes at wave 2.

2. Method

2.1. Participants

A total of 129 children participated in the study. Two thirds were diagnosed with LI ($n=84$), whereas the other one third had a typical development ($n=45$). All children were raised in a monolingual Dutch environment. The TD children were recruited via regular elementary schools and did not have reported language problems. Recruitment of the children with LI was carried out via Royal Dutch Kentalis and Royal Auris Group, which are two national organizations that provide diagnostic, care and educational services for children with communicative problems. In the Netherlands, children are diagnosed with LI if they score at least two SDs below the mean on a standardized language assessment test battery or score at least 1.5 SD below the mean on two out of four predefined subscales that are assessed with at least two appropriate measures. At wave 1, all 84 children in the LI sample met these criteria. At wave 2, seven children did not meet the criteria confirming that the developmental pathways for language are fluid (Reilly et al., 2014). Because these children had a history of LI and research indicates that recovery may be illusory around this age (Scarborough & Dobrich, 1990), we decided not to exclude these children. The children with LI did not have a hearing impairment, intellectual disability or severe articulatory difficulties.

Table 1
Participant characteristics.

	TD	LI	
Girls/Boys	18/27	21/63	$\chi^2(1) = 3.13$, $p = 0.08$
Age in months wave 1 M (SD)	70.78 (8.19)	71.64 (6.64)	$F(1, 127) = 0.42$, $p = 0.52$, $\eta_p^2 = 0.003$
range	59–94	59–87	
Age in months wave 2 M (SD)	82.18 (8.15)	82.92 (6.89)	$F(1, 127) = 0.30$, $p = 0.59$, $\eta_p^2 = 0.002$
range	69–105	70–100	
NVIQ M (SD)	106.98 (15.32)	96.67 (15.54)	$F(1, 127) = 13.03$, $p < 0.001$, $\eta_p^2 = 0.09$
range	81–143	72–131	

Note. TD = typical development; LI = language impairment; NVIQ = NonVerbal IQ. NVIQ scores are standardized scale scores with a mean of 100.

One third of the children with LI attended regular education with ambulatory care ($n = 28$) and two thirds attended special education ($n = 56$).

An overview of the participant characteristics is presented in Table 1. The children with LI and the TD children did not differ in age. They were on average 5;9 (years; months) and 6;9 at wave 1 and wave 2, respectively. The TD group contained relatively more girls than the LI group did, but the difference was not statistically significant. All children had a nonverbal IQ score of 70 or above as measured with the Wechsler Nonverbal Scale of Ability (Wechsler & Naglieri, 2008), showing that no children in the sample had a general learning disorder. The LI group had on average lower nonverbal IQ scores than the TD group, which is typical for studies comparing LI and TD samples (Gallinat & Spaulding, 2014). All children were assessed with a sentence repetition task (a subtest part of the *Taaltoets Alle Kinderen* TAK 'Language test all children', Verhoeven & Vermeer, 2001), which is a language task that is highly sensitive to LI (Conti-Ramsden, Botting, & Faragher, 2001). Within the TD sample 26 children scored above the median (50th percentile) (58%), showing that 50% of the children in the population score lower than they did, while no child with LI fell in this category. A total of 17 TD children (38%) and 17 children with LI (20%) scored below the median, between the 50th and 10th percentile. Two TD children had a score that fell in the lowest 10% of the norm sample (4%), while 64 children with LI fell in this category (76%). From three children with LI, data on the sentence repetition test were missing.

2.2. Measures

2.2.1. Narrative macrostructure

To assess children's macrostructural narrative abilities, the Multilingual Assessment Instrument for Narratives (LITMUS-MAIN; Gagarina et al., 2012; Gagarina, Klop, Tsimpli, & Walters, 2016) was used which includes four comparable stories that are illustrated by six full-color picture sequences. Each story is composed of three episodes for which specific story elements can be expressed. These elements include the introduction of a setting (place and time), a basic event structure (goals, attempts and outcomes) and internal states as initiating event and as reaction. For the purpose of the present study, both narrative comprehension and narrative production were assessed. Children were first presented with a model story (Cat or Dog) followed by ten comprehension questions that are included in the MAIN. These comprehension questions tapped into the goals of the episodes ($n = 3$; e.g., 'Why does the cat jump forwards?'), the internal states of the protagonists ($n = 6$; e.g., 'How does the cat feel?'), and the ability to make inferences about consequences of the events in the story ($n = 1$; e.g., 'Does the boy become friends with the cat? Why?'). A child was awarded one point for each correctly answered comprehension question. Subsequently, children were presented with a different picture sequence (Baby birds or Baby goats) and were asked to generate their own story. Children were awarded one or two points if they introduced their story with a place and/or time (setting). Furthermore, one point was awarded for each expressed internal state as initiating event (e.g., The Baby birds were hungry), goal (e.g., Mother wanted to feed the chicks), attempt (e.g., Mother looked for food), outcome (e.g., Mother came back with a worm), and internal state as reaction (e.g., The Baby birds were satisfied). These five story elements were scored for each of the three episodes and, together with setting, constituted the measure for narrative production, which thus ranged between 0 and 17.

The children were assessed with the same version of the model story and production story at wave 1 and wave 2. Any differences in degree of difficulty between the story versions could not have influenced the results, as there were no significant differences between the TD and LI groups in terms of version of the model story ($\chi^2(1) = 0.014$, $p = 0.90$) or version of the production story ($\chi^2(1) = 0.081$, $p = 0.78$). All narratives were recorded with a highly sensitive microphone (Samson Go Mic) and scored offline by native speakers of Dutch. The interrater reliability for both narrative comprehension and production was good (*see* Boerma et al. (2016) for the exact numbers).

2.2.2. Language measures based on speech sample

The language measures derived from the narratives were lexical diversity (D) and mean length of utterance (MLU) based on the transcriptions of the produced stories. All stories were transcribed using *Codes for the Human Analysis of Transcripts* (CHAT) transcription format, based on the audio recordings. D and MLU were calculated using *Computerized Language Analysis* (CLAN) (MacWhinney, 2000). The narrative task alone did not always yield a sufficient number of utterances per child to be able to calculate reliable measures, especially for children with LI. In addition to the stories, we therefore also transcribed

Table 2
Between-group differences (TD versus LI) on story comprehension and story generation.

	Wave	TD ($n = 45$)	LI ($n = 84$)
Story comprehension (max = 10)	1[a]	9.2 (.9)	8.1 (1.8)
	2[b]	9.3 (.8)	8.9 (1.2)
Story generation (max = 17)	1	8.1 (2.2)	5.2 (1.7)
	2	8.3 (2.2)	7.0 (1.9)

Note. TD = typical development; LI = language impairment
[a] At wave 1, data from two children in the TD sample and two children in the LI sample were excluded because some questions were missing.
[b] At wave 2, data from two children in the LI sample were excluded because some questions were missing.

an informal conversation with the children prior to the narrative task in which the children were asked about a range of accessible topics including their hobbies, birthday and/or favorite tv-show. This conversation allowed us to elicit more utterances, which benefitted the reliability of the language measures. The speech samples (conversation and narratives) were on average 91.5 utterances long. Sample sizes ranged between 39 and 259 utterances.

2.2.3. Standardized language measures

Two standardized measures were used next to the two language measures derived from the narratives. Dutch receptive vocabulary was measured with the Peabody Picture Vocabulary Task (PPVT-III-NL; Schlichting, 2005). The PPVT is a standardized receptive vocabulary test in which a child hears a stimulus word and has to choose the correct referent out of four pictures. The PPVT contains 204 items divided over 17 sets. The sets are ordered according to difficulty and each set consists of twelve items. The PPVT-III-NL was administered and scored according to the official guidelines. This means that the starting set was determined by a child's age and the task was terminated after a child produced nine or more errors within one set. Raw scores were converted to standardized scores based on age-corrected normative scores.

Dutch grammar was measured with the TAK sentence repetition task (Verhoeven & Vermeer, 2001). In this test, each sentence that has to be repeated contains a certain word order property of Dutch and a function word. Only if both the word order and function word in a sentence are repeated correctly, two points are awarded for that particular sentence. If only one of the two is repeated correctly, one point is awarded and if both are repeated incorrectly no point is awarded. The test contained 20 items, resulting in a maximum score of 40. In the analyses, raw scores were used.

2.2.4. Cognitive measures

Verbal short-term memory was measured with a forward digit span task and verbal working memory with a backward digit span task. These tasks were adapted from the Automated Working Memory Assessment (AWMA; Alloway, 2007) and translated into Dutch. In the forward digit span task, children were asked to repeat a sequence of digits in the correct order, assessing the temporary storage of verbal information. The task started with a block of trials with one digit and continued with digit sequences of increasing length, up to seven digits. In the backward digit span task, a similar procedure was followed, with the exception that children were asked to repeat the sequences in backward order and were thus required to both store and manipulate the incoming stimuli. A child proceeded to the next block if four out of six trials were repeated correctly. The task was terminated if three trials within one block were incorrectly repeated. This entailed a wrong ordering of the sequence, an omission of one or more digits or a repetition of one or more incorrect digits. Following the AWMA, a child received one point for each correctly repeated sequence and was awarded six points if the first four trials within a block were correct. If a child scored four out of five items correct this was awarded with five points. Scores could thus range from 0 to 42.

Sustained attention was measured with an integrated visual and auditory Continuous Performance Task (CPT), which was based on the IVA PLUS (Sandford & Turner, 2004). Children were presented with visual and auditory stimuli that could either be a target (number '1′) or a distractor (number '2′). Irrespective of modality, children were asked to press the space bar in response to a target stimulus, but to refrain from responding when a distractor was presented. The test included 168 trials, excluding the practice phase, in which visual and auditory targets ($n = 84$) and distractors ($n = 84$) were randomly interspersedly presented. The task lasted approximately 10 minutes, during which children were required to stay alert and maintain their attention. Overall response sensitivity on this task was scored as d', a statistic from signal detection theory (*see* Macmillan & Creelman, 1991). This inherently dual score reflects percent correct responses to targets (hits) relative to percent incorrect responses to distractors (false alarms). By taking into account both hits and false alarms, this score controls for potential response bias, such as a child pressing the space bar in response to each stimulus. The statistic is calculated as follows: $d' = z(\text{hits}) - z(\text{false alarms})$. The higher the statistic, the better the child's response sensitivity.

2.3. Procedures

The Standing Ethical Assessment Committee of the Faculty of Social and Behavioral Sciences at Utrecht University screened the present study and approved its continuation. An informed consent was signed by parents of the participating children. All children were tested individually in two sessions that each lasted approximately one hour. Testing was done by trained research assistants in a quiet room at a child's school. The test battery included language, memory and attention tasks (some

Table 3
Between-group differences (TD versus LI) for language measures and cognitive measures at wave 1.

	TD ($n = 45$)[a]	LI ($n = 84$)[b]	Statistic
PPVT	111 (13)	96 (12)	$F(1, 126) = 40.91$, $p < 0.001$, $\eta_p^2 = 0.25$
TAK	30 (7)	12 (8)	$F(1, 124) = 158.43$, $p < 0.001$, $\eta_p^2 = 0.56$
D	52 (13)	48 (13)	$F(1, 126) = 2.59$, $p = 0.11$, $\eta_p^2 = 0.02$
MLU	4.8 (.7)	3.7 (.8)	$F(1, 126) = 55.70$, $p < 0.001$, $\eta_p^2 = 0.31$
FDS	21 (4.3)	15 (3.9)	$F(1, 127) = 60.40$, $p < 0.001$, $\eta_p^2 = 0.32$
BDS	12 (3.7)	9 (3.5)	$F(1, 127) = 25.92$, $p < 0.001$, $\eta_p^2 = 0.17$
CPT	2.2 (.8)	1.6 (.7)	$F(1, 126) = 18.30$, $p < 0.001$, $\eta_p^2 = 0.13$

Note. TD = typical development; LI = language impairment; PPVT = Peabody Picture Vocabulary Test = receptive vocabulary; TAK = Taaltoets Alle Kinderen = sentence repetition; D = lexical diversity = expressive vocabulary; MLU = Mean Length of Utterance = grammatical complexity; FDS = Forward Digit Span = verbal short-term memory; BDS = Backward Digit Span = verbal working memory; CPT = Continuous Performance Task = sustained attention.

[a] For one child in the TD sample no digit span data are available.

[b] For three children in the LI sample no TAK data are available.

of which are not reported in this study). The verbal short-term and working memory tasks and the sentence repetition task were the third and fifth task of the first session respectively. Sustained attention and receptive vocabulary were the second and third task of the second session respectively. The second session ended with the narrative task. Children were tested in two waves. For the purpose of the present study, the narrative outcomes of wave 1 and 2 were analyzed while the other measures included in this study were collected at wave 1.

3. Results

3.1. Story comprehension and story generation: TD versus LI

Both groups performed accurately on the model story comprehension questions, with performance of the TD group reaching ceiling. The results are in Table 2.

For story comprehension, an independent samples Mann-Whitney test was performed with group (TD versus LI) as the independent variable because the distribution was skewed to the left. The TD group scored higher than the LI group at wave 1, $U = 1071$, $p < 0.001$, $r = -0.33$. At wave 2 the difference was not significant, $U = 1623$, $p = 0.16$, $r = -0.12$, suggesting an interaction effect. Because the Mann-Whitney test does not allow examining interactions or entering covariates, the analyses were complemented with a mixed-design ANCOVA with group as the between-subjects variable, wave as the within-subjects variable, story comprehension scores as the dependent variable, and nonverbal IQ scores as a covariate. Both group, $F(1,122) = 7.74$, $p < 0.01$, $\eta_p^2 = 0.06$, and nonverbal IQ, $F(1,122) = 4.94$, $p < 0.05$, $\eta_p^2 = 0.04$, were significant, showing better performance for children with TD and children with a higher nonverbal IQ score. The effect of wave was not significant. The interaction between wave and group reached significance, $F(1,122) = 4.19$, $p < 0.05$, $\eta_p^2 = 0.03$, in line with the outcomes of the separate Mann-Whitney tests which showed that only at wave 1 the LI group was outperformed by the TD group; this effect remains significant if a Bonferroni corrected α level of 0.025 is applied.

A mixed-design ANCOVA with group as the between-subjects variable, wave as the within-subjects variable, story generation scores as the dependent variable and nonverbal IQ scores as the covariate showed a significant main effect of group, $F(1,126) = 48.16$, $p < 0.001$, $\eta_p^2 = 0.28$, and a significant interaction effect of group and wave, $F(1,126) = 10.54$, $p < 0.001$, $\eta_p^2 = 0.08$. Neither wave nor nonverbal IQ was significant. Subsequent univariate ANOVA's revealed that both at wave 1, $F(1,126) = 67.82$, $p < 0.001$, $\eta_p^2 = 0.35$, and wave 2, $F(1,127) = 13.71$, $p < 0.001$, $\eta_p^2 = 0.10$, the TD group obtained higher story generation scores than the LI group. The interaction effect is reflected in a difference between the effect sizes showing that the magnitude of the effect of LI becomes smaller as the children grow older.

3.2. Relationships between narrative macrostructure, linguistic and cognitive variables: TD versus LI

Table 3 lists the results of the TD and LI groups on receptive vocabulary, grammar (sentence repetition task), expressive vocabulary (lexical diversity), grammatical complexity (mean length of utterance), verbal short-term and working memory, and sustained attention, all measured at wave 1. The LI group scored lower than the TD group on all measures except expressive vocabulary; the effect sizes are large.

The correlation table (Table 4) lists relationships between narrative macrostructure at wave 1 (story comprehension 1, story generation 1) and wave 2 (story comprehension 2, story generation 2), language abilities (receptive vocabulary, sentence repetition, expressive vocabulary, grammatical complexity) and cognitive abilities (verbal short-term memory, verbal working memory, sustained attention) at wave 1. Thus, only narrative macrostructure was measured at wave 1 and wave 2, all other measures were limited to wave 1. In addition, correlations are listed between age (at wave 1) and all other measures. Correlations that involved story comprehension were calculated using Spearman correlations. In all other cases, Pearson correlations were calculated. The upper right triangle displays the correlations for the TD sample ($n = 45$) whereas the lower left triangle shows the correlations for the LI sample ($n = 84$). Sample size has a larger influence on p-values than

Table 4
Correlations table for the TD sample (upper right triangle) and LI sample (lower left triangle).

	Age1	SC1	SG1	SC2	SG2	PPVT	TAK	D	MLU	FDS	BDS	CPT
Age1		0.04	0.13	0.03	0.36*	−0.06	0.46**	−0.06	0.29	0.43**	0.38**	0.28
SC1	0.29**		0.11	0.17	0.08	0.26	0.08	−0.26	0.07	0.06	0.14	−0.03
SG1	0.31**	0.22		0.21	0.06	0.07	0.27	−0.001	0.43**	0.07	0.27	0.34*
SC2	0.14	0.27*	0.08		−0.04	0.21	0.21	0.01	0.01	0.32*	0.14	0.17
SG2	0.18	0.25*	0.38**			0.004	0.15	−0.11	0.12	0.19	0.12	0.23
PPVT	−0.04	0.30**	0.18	0.24*	0.12		0.45**	−0.10	−0.08	0.53**	0.38*	0.38*
TAK	0.51**	0.35**	0.26*	0.03	0.28*	0.25*		−0.14	0.12	0.67**	0.66**	0.38**
D	0.31**	0.30**	0.20	0.13	0.11	0.08	0.38**		0.19	−0.06	−0.08	−0.27
MLU	0.38**	0.26*	0.37**	−0.03	0.25*	0.08	0.58**	0.48**		0.01	0.33*	0.22
FDS	0.40**	0.30**	0.20	0.17	0.28**	0.22*	0.55**	0.22*	0.33**		0.60**	0.51**
BDS	0.40**	0.36**	0.23*	0.13	0.38**	0.12	0.39**	0.31**	0.36**	0.57**		0.50**
CPT	0.15	0.27*	0.31**	0.22*	0.36**	0.21	0.30**	0.07	0.24*	0.38**	0.37**	

Note. SC1 = Story Comprehension wave 1; SG1 = Story Generation wave 1; SC2 = Story Comprehension wave 2; SG2 = Story Generation Wave 2; PPVT = Peabody Picture Vocabulary Test = receptive vocabulary; TAK = Taaltest Alle Kinderen = sentence repetition; D = lexical diversity = expressive vocabulary; MLU = Mean Length of Utterance = grammatical complexity; FDS = Forward Digit Span = verbal short-term memory; BDS = Backward Digit Span = verbal working memory; CPT = Continuous Performance Task = sustained attention; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

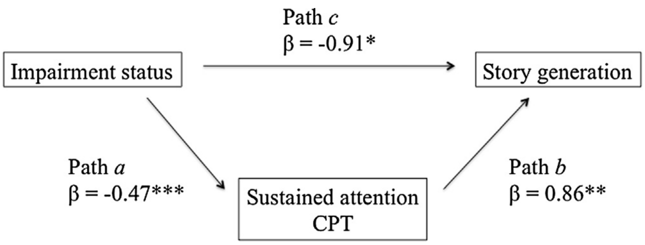

Fig. 1. Mediation model with impairment status as the independent variable, sustained attention as the mediator variable and story generation as the dependent variable; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

on effect sizes, which may explain why some effect sizes are larger for TD than LI but are only statistically significant in the LI sample.

In the TD sample, story comprehension at wave 2 correlated significantly with verbal short-term memory, while story generation at wave 1 was correlated with grammatical complexity and sustained attention. In the LI sample, age at wave 1 was significantly correlated with both story comprehension and generation ate wave 1, all language and cognitive measures were significantly correlated with story comprehension at wave 1, while at wave 2, receptive vocabulary and sustained attention correlated significantly with story comprehension. Sentence repetition, grammatical complexity, verbal working memory and sustained attention were significantly correlated with story generation at wave 1. Sentence repetition, mean length of utterance, verbal short-term memory, verbal working memory and sustained attention correlated significantly with story generation at wave 2.

Next, a mediation analysis was performed. The skewed comprehension outcomes did not allow linear regression analyses and this analysis was therefore performed on the story generation outcomes only. The story generation scores at wave 2 were the dependent variable, receptive vocabulary, sentence repetition, expressive vocabulary, grammatical complexity, verbal short-term memory, verbal working memory and sustained attention at wave 1 were mediator variables (M1 = PPVT, M2 = TAK, M3 = D, M4 = MLU, M5 = FDS, M6 = BDS, M7 = CPT), and impairment status the independent variable. The distinction between TD and LI was based on the diagnosis of LI, which preceded wave 1. A multiple mediation analysis was applied, using the PROCESS application for SPSS (Hayes, 2013). Path *a* refers to the indirect effect of the independent variable (impairment status) on the mediators. Path *b* refers to the indirect effect of the mediators (receptive vocabulary, sentence repetition, expressive vocabulary, grammatical complexity, verbal short-term memory, verbal working memory, sustained attention) on the dependent variable (story generation). Path *c* refers to the direct effect of the independent variable (impairment status) on the dependent variable (story generation). According to the well-known requirements for mediation as specified by Baron and Kenny (1986), mediation requires all paths to be statistically significant. Nonverbal IQ was added as a covariate, to be better able to detect causal relationships.

It turned out that for path *b* only sustained attention was significant – the regressions are in Appendix B – and we continued with a mediation model with one mediator variable (M1 = sustained attention), illustrated in Fig. 1. The path from impairment status to sustained attention (*a* path) was statistically significant, as was the path from sustained attention to story generation (*b* path). Finally, the direct effect of the independent variable, impairment status, on the dependent variable, story generation was significant (*c* path). To determine the significance of indirect effects bootstrapped tests and confidence intervals are more reliable than *p*-values. These results confirm that the indirect effect of sustained attention on story generation, 95% CI [−0.86, −0.14], is significant; the number of bootstraps was set at 5,000 (bias-corrected). The unstandardized indirect effect (i.e., the product of the *a* and *b* coefficients) is −0.40. This implies that the score on story generation is expected to decrease by 0.40 units (on a scale from 0 to 17) if a child has LI compared to TD if one considers only the indirect influence via sustained attention.

4. Discussion and conclusions

For this study, it was investigated if children with LI perform lower than TD children on narrative macrostructure measures in a comprehension and generation task and if these effects were also found one year later. The children first listened to a story told by the experimenter (model story), followed by ten comprehension questions. Then, their task was to generate a story based on six pictures. This story was different from the model story but similar in complexity. The inclusion of multiple linguistic and cognitive measures and the two-wave design allowed examining the cause of any between-group differences between TD and LI on narrative macrostructure, and determining if linguistic impairments, cognitive limitations or both types of factors explain the lower performance of children with LI on narrative macrostructure.

Story comprehension was evaluated with ten comprehension questions about goals, internal states and inferences. The TD group had higher comprehension scores than the LI group at wave 1, when the children were five and six year old. Story generation was evaluated based on aggregated measures that contained information on the setting, basic event structure (goals, attempts and outcomes), and internal states of the protagonists at the beginning and end of the event. At both time points, the children with TD generated more complex stories than the children with LI did. These findings confirm that LI is associated with lower performance on narrative macrostructure (Bishop & Donlan, 2005; Dodwell & Bavin, 2008; Gillam & Carlile, 1997; Manhardt & Rescorla, 2002; Merritt & Liles, 1987; Miranda et al., 1998; Reilly et al., 2004; Wright & Newhoff, 2001). Most research on LI has focused on the children's expressive language abilities (Andreu et al., 2011). The lower performance on narrative comprehension shows that LI also affects children's receptive abilities, in line with findings reported in other studies (Bishop & Adams, 1992; Dodwell & Bavin, 2008; Merritt & Liles, 1987).

Not all previous studies observed effects of LI on narrative macrostructure (Iluz-Cohen & Walters, 2012; Liles et al., 1995; Merritt & Liles, 1987; Norbury & Bishop, 2003) and type of task could be one of the factors that may explain the conflicting findings (Duinmeijer et al., 2012; Merritt & Liles, 1989). In the present study the discrepancy between comprehension and generation may support this idea, assuming that the generation task, in which children produce a narrative, is more demanding than the comprehension task. According to Levelt's model (Levelt, 1999), production involves conceptualization of the pre-verbal message, formulation (selecting and sequencing of lexical items), articulation and monitoring. In particular in narrative tasks, which involve conceptualizing links between events, conceptualization may take up a large amount of a child's attentional resources. This may explain not only the role of sustained attention, as will be discussed later, but also why children tend to perform better on the comprehension than on the generation task.

Both the analyses of story comprehension and generation revealed a significant interaction between wave and group, suggesting that the gap in narrative macrostructure performance between TD and LI narrows because the LI group develops faster than the TD group. However, in comprehension, only the LI group could show growth because the TD group performed at ceiling at wave 1. Consequently, the generation scores are more reliable for comparing development in the two groups. The children with LI went to special education or received ambulatory care and the greater growth in the LI group could be influenced by the treatment they receive. Based on comparisons of the growth trajectories of children with TD and LI in the domain of vocabulary and grammar, Rice (2013) concluded that children with LI tend to show a delayed onset of development but that their developmental trajectories resemble those of their peers with TD. If the same holds for narrative development, younger children with TD should show the same steep growth as the LI group in this study. In the present study, data from younger children were not available but future research could compare narrative development across children with LI and younger TD peers. Taken together, these findings suggest that it could be possible that age and task demands could explain why some studies find effects of LI on narrative macrostructure, whereas others do not.

The second goal of the study was to investigate which factors cause the weaker performance of children with LI on narrative macrostructure. A multiple mediation analysis revealed that sustained attention, which is the ability to keep focused, predicted story generation outcomes (path *b*), in line with findings by Duinmeijer et al. (2012). Sustained attention was weaker in the LI than in the TD sample (path *a*), as were story generation outcomes at wave 2 (path *c*), hence all criteria for mediation were met. The observation that sustained attention mediated the relationship between LI and narrative and story generation underscores Ebert and Kohnert's (2011) conclusion that research should consider the role of sustained attention weaknesses in LI. In contrast with other research (Dodwell & Bavin, 2008; Duinmeijer et al., 2012), verbal memory did not play a role. In these studies, effects only emerged for a word recall task as a measure of verbal memory, and not when a digit span task was used. In addition, narrative performance was tested with a retell task, which may explain different findings across studies. Linguistic variables did not predict narrative macrostructure, which confirms that narrative quality at a global level is relatively insensitive to language level (Boerma et al., 2016; Hipfner-Boucher et al., 2014; Pearson, 2002).

There are a number of limitations to our study and directions for future research that are relevant to mention. The left skew in the comprehension outcomes points to ceiling performance which prevented us, first, from observing growth in the TD group and, second, from performing a mediation analysis. The correlation patterns suggest that the story generation outcomes cannot be generalized to story comprehension, e.g., receptive vocabulary may play a more substantial role in story comprehension than story generation. The generalizability of the story generation findings is furthermore limited by type of task. Possibly, verbal memory does mediate relationships between LI and narrative macrostructure when a narrative retell task is used. In future research it would be important to administer a more demanding comprehension task and to investigate task effects, ideally comparing story generation with and without a model story and story retell. Finally, the present study revealed that age influences LI children's performance on narrative macrostructure, but to enhance our insight into the developmental trajectories of the narrative skills of children with LI and to more reliably investigate causal relationships with linguistic and cognitive skills, longitudinal research with three or more time points is needed.

The present study demonstrates that children with LI have difficulties with narrative macrostructure, which refers to the analysis of narratives at a global level. Effects of LI emerged for both story comprehension and story generation, but the gap between TD and LI narrowed over time. Weak sustained attention, and not limited verbal memory or language abilities, explained why children with TD performed better on narrative macrostructure than children with LI, confirming the relevance of sustained attention for the verbal behavior of children with LI, particularly in tasks with extended discourse, such as narrative tasks. The results of this study have important implications for understanding the profile of children with LI and suggest causal links between nonverbal cognitive abilities and performance on verbal tasks. The outcomes of the mediation analysis suggest that interventions that aim at improving children's narrative skills at a global level should focus on children's ability to keep focused.

Acknowledgements

This work is part of the research program 'Cognitive development in the context of emerging bilingualism: Cultural minority children in the Netherlands' which is financed by a VIDI grant awarded to dr. Elma Blom (PI) by the Netherlands Organization for Scientific Research (NWO). We thank the children, parents and schools that participated in the study.

Appendix A. : Examples of Baby goats story TD and LI (CHAT conventions removed)

Boy TD (88 months)

Er waren twee geitjes. En er was een geitje in het water gevallen. En toen kwam zijn vader aangerend en die zag hem. En die hielp hem uit het water. En er was ook nog een wolf en die zag dat andere geitje. En die wou hem opeten. En toen sprong die naar voren. En toen zag dat geitje hem. En toen schrok die. En toen pakte hij hem bij z'n been. En toen was er ook nog een vogel. En hij zag hem. En die wou dat geitje helpen. En toen beet ie hem in ze staart. En toen jaagt ie hem weg. En toen was z'n vader blij dat ie ze geitjes weer terug had.

There were two little goats. And one little goat had fallen in the water. And then his father came running and saw him. And he helped him out of the water. And there was also a wolf and he saw the other little goat. And he wanted to eat him. And then he leaped forward. And then the little goat saw him. And then he was startled. And then he grabbed him at the leg. And then there was also a bird. And he saw him. And he wanted to help that little goat. And then he bit him in his tail. And then he scared him away. And then his father was happy that he got his little goats back.

Boy LI (86 months)

De lammetjes gingen weg. Toen ging hij in een water. Toen ging de moeder even eruit. En ze wist niet dat de vos daar even wegkwam. En toen de vos het zag toen ging hij het lammetje pakken. Toen kwam de meeuw eraan. Die beet in ze staart. En toen ging de vos weg.

The little lambs went away. Then he went into a water. Then his mother went out. And she did not know that the wolf was just coming away. And when the fox saw it he grabbed the little lamb. Then the seagull came. He bit in his tail. And then the fox went away.

Appendix B. : Regressions path *b* model with all mediators included

	Coefficient	Standard error	*t*-statistic	*p*-value
Constant	5.87	2.40	2.45	0.02
PPVT	−0.004	0.02	−0.29	0.78
TAK	0.02	0.03	0.63	0.53
D	0.001	0.01	0.09	0.93
MLU	0.16	0.26	0.62	0.54
FDS	0.02	0.06	0.30	0.77
BDS	0.07	0.06	1.07	0.29
CPT	0.68	0.27	2.56	0.01
Impairment status	−0.09	0.54	−0.17	0.86
NVIQ	−0.01	0.01	0.94	0.35

Note: PPVT = Peabody Picture Vocabulary Test = receptive vocabulary; TAK = Taaltest Alle Kinderen = sentence repetition; D = lexical diversity = expressive vocabulary; MLU = Mean Length of Utterance = grammatical complexity; FDS = Forward Digit Span = verbal short-term memory; BDS = Backward Digit Span = verbal working memory; CPT = Continuous Performance Task = sustained attention; NVIQ = nonverbal IQ.

References

Alloway, T. P. (2007). *Automated Working Memory Assessment (AWMA)*. London: Pearson.

Andreu, L., Sanz-Torrent, M., Guàrdia Olmos, J., & MacWhinney, B. (2011). Narrative comprehension and production in children with SLI: an eye movement study. *Clinical Linguistics & Phonetics*, *25*(9), 767–783.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic: and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bishop, D. V. M. (2010). Which neurodevelopmental disorders get researched and why? *PLoS ONE*, *5*(11).

Bishop, D. V. M., & Adams, C. (1992). Comprehension problems in children with specific language impairment: literal and inferential meaning. *Journal of Speech and Hearing Research*, *35*(1), 119–129.

Bishop, D. V. M., & Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: evidence from specific language impairment. *British Journal of Developmental Psychology*, *23*(1), 25–46.

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, *52*, 156–173.

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: implications for clinical practice. *International Journal of Language and Communication Disorders*, http://dx.doi.org/10.1111/1460-6984.12234

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, *18*(1), 1–21.

Boudreau, D. (2008). Narrative abilities: advances in research and implications for clinical practice. *Topics in Language Disorders*, *28*, 99–114.

Boudreau, D. M., & Hedberg, N. L. (1999). A comparison of early literacy skills in children with specific language impairment and their typically developing peers. *American Journal of Speech-Language Pathology*, *8*(3), 249–260.

Clahsen, H. (1989). The grammatical characterization of developmental dysphasia. *Linguistics*, *27*, 897–920.

Cleave, P. L., Girolametto, L. E., Chen, X., & Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning children with specific language impairment. *Journal of Communication Disorders*, *43*(6), 511–522.

Colozzo, P., Gillam, R. B., Wood, M., Schnell, R. D., & Johnston, J. R. (2011). Content and form interactions in the narratives of children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *54*, 1609–1627.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748.

De Villiers, J. (2000). Language and theory of mind: what are the developmental relationships. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: perspectives from developmental cognitive neuroscience, Second Edition* (pp. 83–123). Oxford: Oxford University Press.

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: an investigation of their narratives and memory. *International Journal of Language and Communication Disorders*, *43*(2), 201–218.

Duinmeijer, I., de Jong, J., & Scheper, A. R. (2012). Narrative abilities and memory and attention in children with a specific language impairment. *International Journal of Language & Communication Disorders*, *47*(5), 542–555.

Ebert, K. D., & Kohnert, K. (2011). Sustained attention in children with primary language impairment: a meta-analysis. *Journal of Speech, Language, and Hearing Research*, *54*, 1372–1384.

Epstein, S., & Philips, J. (2009). Storytelling skills of children with specific language impairment. *Child Language Teaching and Therapy*, *25*, 285–300.

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: a study of SLI and normal development. *Journal of Child Language*, *31*, 661–681.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Valimaa, T., Balčiūnienė, I., & Walters, J. (2012). MAIN −Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, *56*, 1–155.

Gagarina, N., Klop, D., Tsimpli, I., & Walters, J. (2016). Introduction to the special issue on narrative abilities in bilingual children. *Applied Psycholinguistics*, *37*(1), 11–17.

Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: a meta-analysis. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1363–1382.

Gillam, R. B., & Carlile, R. M. (1997). Oral reading and story retelling of students with specific language impairment Language. *Speech & Hearing Services in the Schools*, *28*(1), 30–42.

Gillam, R. B., & Johnston, J. R. (1992). Spoken and written language relationships in language/learning-impaired and normally achieving school-age children. *Journal of Speech and Hearing Research*, *35*(6), 1303–1315.

Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis: a regression-based approach*. New York, NY: Guilford Press.

Heilmann, J., Miller, J., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology*, *19*, 154–166.

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

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2014). Narrative abilities in subgroups of English language learners and monolingual peers. *International Journal of Bilingualism*, 1–16.

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: narratives of preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, *15*(1), 58–74.

Leonard, L. B. (1995). Functional categories in the grammars of children with Specific Language Impairment. *Journal of Speech and Hearing Research*, *38*(6), 1270–1283.

Leonard, L. (2014). *Children with specific language impairment* (2nd ed.). Cambridge: MIT Press.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*, 223–232.

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research*, *38*, 415–425.

Liles, B. Z. (1993). Narrative discourse in children with language disorders and children with normal language: a critical review of the literature. *Journal of Speech and Hearing Research*, *36*(5), 868–882.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk.* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: a user's guide*. Cambridge: Cambridge University Press.

Manhardt, J., & Rescorla, L. (2002). Oral narrative skills of late talkers at ages 8 and 9. *Applied Psycholinguistics*, *23*, 1–21.

Merritt, D. D., & Liles, B. Z. (1987). Story grammar ability in children with and without language disorder: story generation, story retelling, and story comprehension. *Journal of Speech and Hearing Research*, *30*, 539–552.

Merritt, D. D., & Liles, B. Z. (1989). Narrative analysis: clinical applications of story generation and story retelling. *Journal of Speech and Hearing Research*, *54*, 429–438.

Miranda, A. E., McCabe, A., & Bliss, L. S. (1998). Jumping around and leaving things out: a profile of the narrative abilities of children with specific language impairment. *Applied Psycholinguistics*, *19*, 647–667.

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and specific language impairment: an update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology*, *19*, 78–94.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287–313.

Norbury, C. F., Gemmel, T., & Paul, R. (2014). Pragmatics abilities in narrative production: a cross-disorder comparison. *Journal of Child Language*, *41*(3), 485–510.

O'Neill, D. K., Pearce, M. J., & Pick, J. L. (2004). Preschool children's narratives and performance on the Peabody Individualized Achievement Test–Revised: evidence of a relation between early narrative and later mathematical ability. *First Language*, *24*(71), 149–183.

Owens, R. E. (1996). *Language development*. Boston, MA: Allyn & Bacon.

Paul, R., & Smith, R. (1993). Narrative skills in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech, Language, and Hearing Research*, *36*, 592–598.

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller, & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Clevedon, UK: Multilingual Matters.

Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). Frog, where are you? Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language*, *88*(2), 229–247.

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., & Wake, M. (2014). Specific language impairment: a convenient label for whom? *International Journal of Language & Communication Disorders*, *49*(4), 416–451.

Rice, M. (2013). Language growth and genetics of specific language impairment. *International Journal of Speech-Language Pathology*, *15*(3), 223–233.

Rice, M., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: a longitudinal study from 2;6 to 21 years of age. *Journal of Speech Language and Hearing Research*, *58*(2), 345–359.

Rice, M., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in english-Speaking children. *Journal Speech and Hearing Research*, *39*(6), 1239–1257.

Sandford, J. A., & Turner, A. (2004). *IVA + plus: integrated visual and auditory continuous performance test administration manual*. Richmond, VA: BrainTrain.

Scarborough, H., & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech, Language, and Hearing Research*, *33*(1), 70–83.

Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL. Dutch version*. Amsterdam: Harcourt Assessment B.V.

Soodla, P., & Kikas, E. (2010). Macrostructure in the narratives of estonian children with typical development and language impairment. *Journal of Speech, Language, and Hearing Research*, *53*(5), 1321–1333.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing II*. Norwood, N.J: Ablex.

Tabors, P., Snow, C., & Dickinson, D. (2001). Homes and schools together: supporting language and literacy development. In D. Dickinson, & P. Tabors (Eds.), *Beginning literacy with language* (pp. 313–334). Baltimore, MD: Brookes.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech Language and Hearing Research*, *40*, 1245–1260.

Uccelli, P., & Paíez, M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, *38*, 225–236.

Verhoeven, L., & Vermeer, A. (2001). *Taaltoets Alle Kinderen (TAK)*. Arnhem: Cito.

Wechsler, D., & Naglieri, J.A. (2008). Wechsler Nonverbal Scale of Ability. Purchased from: http://www.pearsonclinical.nl/wnv-nl-wechsler-non-verbal.

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

Wright, H. H., & Newhoff, M. (2001). Narration abilities of children with language-learning disabilities in response to oral and written stimuli. *American Journal of Speech-Language Pathology*, *10*(3), 308–319.