

Identifying early preschool bilinguals at risk of DLD: A composite profile of narrative and sentence repetition skills¹

Natalia Gagarina

Leibniz-Centre General Linguistics (ZAS), Berlin, Germany

Stefanie Gey

Leibniz-Centre General Linguistics (ZAS), Berlin, Germany

Natalie Sürmeli

Leibniz-Centre General Linguistics (ZAS), Berlin, Germany

For this study one hundred sixty-seven Russian-/Turkish-German preschool children were tested with a battery of language proficiency tests in both languages. On the basis of 1.5 SD below monolingual norm for L2 German and 1.25 SD below bilingual mean for either home language, 9 children at risk of developmental language disorders (DLD) (mean age of 4 years and 5 months) were identified and 16 age-matched TD children were selected out of the cohort. All these children were tested with the LITMUS-MAIN and –SR tests in German. The results across TD and at risk of DLD group were compared. TD clearly outperformed at risk of DLD in SR. In elicited narratives, macrostructure and microstructure were scrutinized across groups. Similar to the previous findings, our results show significant differences between at risk of DLD und TD in the microstructure, e.g. total number of word tokens and verb-based communication units and SR. For the macrostructure, TD outperformed at risk children only for story complexity. The study expands our knowledge on the cut-off criteria for the identification of bilinguals at risk of DLD, scrutinized very early narratives for bilinguals at risk of DLD features and questions the similarity of cognitive skills in TD and at risk of DLD children.

1 Introduction and background

“Jumping around and leaving things out” wrote Miranda, McCabe & Bliss (1998) about the narratives of monolingual English-speaking children with

¹ The present study was supported in part by Bundesministerium für Bildung und Forschung (BMBF), Grant No. 01UG0711 and by the Berliner Senate.

developmental language disorders (henceforth, DLD). The same is true for the bilingual DLD population, whose narrative production is much less investigated (e.g. Tsimpli et al., 2016, Altman et al., 2016) despite their larger number than monolinguals. Due to migration, the number of children not speaking the societal language at home has dramatically increased (WMO report, 2013) and in some major cities' districts, like Berlin's district of Neukölln, they compose about 70% of the entire population (Bettge & Oberwöhrmann, 2012:15). Some of these children show an atypical language acquisition pattern which – on the surface – is similar to monolingual DLD. These children are examined with tests normed for monolingual children and are therefore often misdiagnosed. With the aim of making adequate access to bilinguals with language problems possible, researchers in the COST Action IS0804 created the Language Impairment Testing in Multilingual Settings (LITMUS) test battery (Armon-Lotem, Meir, & de Jong, 2015). This study applies two tests out of this battery: *Multilingual Assessment Instrument for Narratives* (LITMUS-MAIN) (Gagarina et al., 2012) and SRep task for German (Marinis & Armon-Lotem, 2015; For German: Hamann, Chilla, Ruigendijk, & Abed Ibrahim, 2013) for examining the preschool Russian-/Turkish-German bilingual's language production in German. These tests were selected out of the bigger cohort of tests on the basis of the cut-off criteria described below.

The choice of narratives is justified by three overarching arguments. First, not only do narrative skills compose the basics of communication, they are fundamental for children's literacy, reading and even life-long success (e.g. Bishop & Edmundson, 1987; Bliss, McCabe, & Miranda, 1998; Gutiérrez-Clellen, 2002; Hayward & Schneider, 2000; McCabe, 1996; McCabe & Rollins, 1994; Norris & Bruning, 1988; Swanson, Mills, Hood, & Fey, 2005; Torrance & Olson, 1984; Wallach, 2008). Second, narrative tests have proved to be indispensable in clinical linguistics (e.g. Heilmann, Miller, Nockerts, & Dunaway, 2010) for diagnostic specific language impairment (Hayward & Schneider 2000; Ringmann & Siegmüller, 2013; Schneider, Hayward & Dube, 2006) and are not strongly biased towards one of the dual languages of a bilingual child (Paradis, Genesse & Crago, 2010). Previous studies showed robust differences between TD vs. DLD (monolingual and bilingual) narrative's typical features. These are microstructural features, such as story length, lexical diversity (e.g. Altman et al., 2016; Iluz-Cohen & Walters, 2011; Tsimpli et al., 2016) as well as the variation in somewhat controversial features of macrostructure across these two populations (Iluz-Cohen & Walters, 2011, but Squires et al., 2014). Finally, elicited narratives are semi-spontaneous data and allow us, by looking into their macro- and microstructure organization, to evaluate both cognitive and linguistic skills of bilinguals (Liles, 1993).

DLD is determined here after Leonard (1998) as a language performance disorder that appears alongside typical intellectual, neuro-psychological and general cognitive development (cf. Paradis, Genesee & Crago, 2011).

2 Theoretical background

2.1 Narratives: Macrostructure and Microstructure

Narratives can be said to be organized at two levels: the macrostructure and microstructure level (Liles, Duffy, Merritt & Purcell, 1995). Macrostructure is the global organization of content which is traditionally described by *story grammar model* (Mandler, 1979; Stein & Glenn, 1979). According to Stein and Glenn (1979), story grammar includes the setting (which introduces the characters and describes the non-linguistic context of a story), and the following components of an episode: an initiating event, internal responses, internal plans, attempts, direct consequences and reactions (Peterson & McCabe, 1991; Stein & Glenn, 1979). These combined components in a narrative are referred to as *story structure*. Additionally, the combination of story structure components, which comprise a full complex episode, is a crucial part of the macrostructure evaluation. A full complex episode consists of the *Goal* of the protagonist, the *Attempt* to reach the Goal and, finally, the *Outcome* of the protagonist's actions. Not all components of a full complex episode are realized in the production of narratives, especially in children's narratives. In order to capture different levels of complexity, Westby (2005) suggested a binary decision tree, which evaluates: i) sequences with no Goal statement, ii) incomplete episodes, with a Goal statement but no Goal-Attempt-Outcome structure (henceforth, GAO), and iii) the most complex and complete episodes which include all three GAO components (McCabe & Peterson, 1984; Trabasso & Nickels, 1992). The GAO represents the highest level of story complexity and is an indicator of the ability to produce a coherent story. So, the basic analysis of narrative macrostructure adopted in the present article includes a quantitative estimate of a narrative's macrostructure, which is the number of episode components (i.e., story structure) and a qualitative evaluation capturing the combination of the episode components (i.e., story complexity).

At the interface between macrostructure and microstructure, lies the so-called mental state language, which can be grasped through Internal State Terms or Internal States (henceforth, IS) and is indispensable for the evaluation of narrative skills. IS reflect Theory of Mind properties (Curenton & Justice, 2004) and show children's skills to build inferences (Nippold, Ward-Lonergan & Fanning, 2005; Westby, 2005). As proposed in LITMUS-MAIN (Gagarina et al., 2012; 2015), IS include perceptual state terms, for example, *see, hear, feel, smell*; physiological state terms, for example, *thirsty, hungry, tired, sore*; consciousness

terms, for example, *alive, awake, asleep*; emotion terms, such as *sad, happy, angry, worried, disappointed*; mental verbs, including *want, think, know, forget, decide, believe, wonder, have/ make a plan*; linguistic verbs/ verbs of saying/ telling, for example, *say, call, shout, warn, ask*.

Microstructure is known to be more language-specific and it includes length as measured in e.g., total number of tokens, types, lexemes or communication units (see below) and lexis, which in turn can be evaluated in terms of diversity of different word classes or different items within one word class. Microstructure also covers language-specific items, which are responsible for morphosyntax and coherent discourse composition.

The present study scrutinizes both these levels of narratives: macrostructure and microstructure.

2.2 Narrative skills (Macrostructure and Microstructure) in bilingual TD and DLD children

While there are vast amounts of studies comparing narrative skills in monolingual typically-developing (henceforth, TD) and DLD children (e.g. Liles, 1987; Merrit & Liles, 1987; Skerra, Adani, & Gagarina, 2013), only a few exist that investigate macrostructure and microstructure in bilingual TD as compared to DLD. These include, to our knowledge, two studies on English-Hebrew speaking children (Iluz-Cohen & Walters, 2011; Altman et al., 2016), an investigation on bilinguals with different home languages who speak Greek, i.e. L1 (Tsimpli et al., 2016) and several investigations into Spanish-English speaking children (Cleave, Girolametto, Chen & Johnson, 2010; Gutiérrez-Clellen, Simon-Cerejido & Leone, 2009; Simon-Cerejido & Gutiérrez-Clellen, 2009; Squires et al., 2014) at preschool and primary school age.

The studies examining narrative skills in the similar – as far as age is concerned – population and using comparable (to the present study) evaluation of macrostructure and microstructure are reviewed. Iluz-Cohen & Walters (2011) applied the story grammar model to the investigation of narrative macro- and microstructure in preschool English-Hebrew speaking bilinguals. 8 TD and 9 DLD children were asked to tell a story after having looked at a picture book familiar to them. While no differences were found for the story structure, DLD scored significantly lower on the microstructure measures of lexis and morphosyntax. In another study, Altman et al. (2016) used LITMUS-MAIN (for details see section Method) to elicit retold narratives from 12 DLD and 19 age-matched TD English-Hebrew preschoolers. For the macrostructure analysis, Altman et al. employed the story grammar model (e.g. Mandler, 1979; Stein & Glenn, 1979), which he combined with the binary decision tree (Westby, 2005) to evaluate the levels of story complexity. The analysis of microstructure targeted

story length, lexis, and morphosyntax. Their results corroborate previous findings on comparable macrostructure production in TD and DLD in both their languages. Similar to previous findings, the elements of microstructure, such as verb production, length of stories measured in communication units, lexical diversity was different in TD and DLD children.

Tsimpli et al. (2016) dealt with L2 Greek in 5 to 11-year-old bilingual children in retold narratives elicited with LITMUS-MAIN. Macro- and microstructure of 15 DLD and 15 age-matched TD bilinguals were compared to the two age-matched groups of TD and DLD Greek-speaking monolinguals. The results confirmed the impoverished performance of bilingual DLD in—more language-specific—microstructure, including lexical diversity and internal state terms. Bilingual TD and DLD children did not differ in terms of macrostructure, however bilingual DLD outperformed monolingual DLD on measures of story structure.

Squires et al. (2014) found poorer macro- and microstructure performance for 21 DLD Spanish-English preschoolers at mean age 5 years 9 months as compared to TD age-matched children.

The findings of Skerra et al. (2013), although her study compared monolingual German-speaking TD and DLD children, is especially relevant for ours. Skerra and colleagues documented significant differences in the macrostructure of the elicited LITMUS-MAIN narratives between TD and DLD in the amount of full episodes whose components were tied together by connectors. For example, DLD and TD children differed significantly in the average scores for story complexity. The difference was however highly significant ($p=.007$) between two groups in GA episode with no connection between episode components *Die Katze will den Schmetterling. Sie springt da.* ‘The cat want the batterfly. It jumps there.’ vs. GA episode with syntactically connected components *Deswegen ist die Mutter weggegangen und hat was zu fressen geholt* ‘That’s why the mother flew away and got something to eat’ (examples from Skerra et al., 2013:7-8). In particular, they found, that the number of Attempt-Outcomes is identical between TD and DLD, whereas syntactically connected Attempt-Outcome are produced significantly more frequently by TD. They conclude that it is not the pure macrostructure (i.e. story complexity) with its different levels that distinguish TD and DLD, but a tandem of macrostructural and microstructural components which should be treated together.

In sum, previous findings show no consensus on whether macrostructure differentiates bilingual TD and DLD children. Similar findings to ours reveal that bilingual TD and DLD clearly differ in length of narratives, lexical diversity, and morpho-syntax. Our study builds upon these findings and examines in detail the different measures of narrative length and the *grammatical quality*, i.e. the presence of verbs in the communication units.

2.3 Sentence Repetition (SR)

While narrative macrostructure was only recently used and is still debated as a DLD marker in bilinguals, there seems to be “a long tradition of using SRep tasks <...> to measure language abilities in a speech and language therapy setting (Conti-Ramsden, Botting, & Faragher, 2001) with clinical assessments of language often including a sentence recall subtest“ (Marinis & Armon-Lotem, 2015:97). Within COST Action IS0804, SR was developed within the framework of the Language Impairment Testing in Multilingual Settings (LITMUS) test battery (Armon-Lotem et al., 2015) for bilingual children in almost two dozen languages. Growing research on Sentence Repetition in bilingual acquisition showed that SR can be an accurate cue for identifying bilingual DLD (Thordardottir & Brandeker 2013, Chiat et al., 2013). For example, Meir et al. (2015) found “four main error patterns which distinguish children with biSLI from those with biTLD: sentence fragments, omission of coordinators and subordinators, omission of prepositions, and simplification of *wh*-questions and relative clauses” in their study with 30 TD and 15 DLD Russian-Hebrew sequential bilinguals (Meir et al., 2015:442).

For L2 German, the study of Lomako, Held, Lapenko, Gagarina & Lindner (2015) performed a detailed analysis to find early markers that can identify bilinguals at risk of DLD in a longitudinal study with more than 100 preschool children. This study used a sentence repetition test for German with 30 sentences, covering three levels: Level I: SV, SVO, SVO-particle verbs, SVO-modal verbs, Level II: Negation, WH-Questions, Topicalization, SVOO and Level III: bi-clausal structures: coordinated and sub-ordinated sentences (adverbial-, object-, relative clauses). In the sentences, only present and past perfect tense, early acquired nouns & verbs and nouns mostly in singular with definite articles were used. They found that children at risk more often omit determiners and use the unspecified form **de*, an equivalent to ‘the’ (**de Frau* ‘the woman’, **de Mann* ‘the man’), than TD children.

3 The Study

The present study expands research on bilingual narratives to two language pairs: Russian-German and Turkish-German. In doing so, it addresses the very onset of the development of narrative skills in German in two groups of simultaneous bilinguals with a mean age of 4 years and 5 months and a mean age of onset of 10 months. The two groups – children at risk of DLD and with a matched socio-economic status (SES) and age-matched group of controls – were selected from a bigger cohort of children during the following process. First, one hundred sixty-

seven preschoolers were tested several times within an interval of about 9 months using a battery of tests (see section method). On the basis of their dual language performance in proficiency tasks in the first testing, and (where available) in the second testing, the at-risk group was identified, and the age-matched group was selected. In order to ensure the accuracy of the selection, SR test was performed (cf. Marinis & Armon-Lotem, 2015 on the accuracy of SR task for the identification of bilingual DLD) and the background criteria, such as age of onset, length of exposure, etc. were controlled for. After this process, 9 at risk of DLD and 16 TD children were left in the study.

For these two groups, we first compared the two components of macrostructure: story structure and story complexity. Previous comparisons of TD and DLD provided contradictory results: while some studies found significant differences between TD and DLD in macrostructure (e.g. Squires et al., 2014), the others did not (Iluz-Cohen & Walters, 2011; Altman, Armon-Lotem, Fischman & Walters, 2016). Given the early age of children – at which productive narrative skills are about to develop – we expect no significant differences between the two groups. The second analysis targeted the production of IS: the total number of IS tokens as well as the number of different lexemes was counted. Additionally, the percentage (number of IS normed by story length) was calculated. We expect similarities to the previous findings for IS in bilinguals, which are at the interface between macro- and microstructure (Altman et al., 2016), as well as for the other measures of microstructure, such as story length in tokens and types, and the number of verb-based Communication Units (henceforth, CU). We furthermore expect to see significant differences between the two groups (cf. Iluz-Cohen & Walters, 2011; Squires et al., 2014; Tsimpli et al., 2016).

It is microstructure that is expected to be one of the main cues for the differentiation of DLD and TD when their narrative skills are still rudimentary. For macrostructure, we predicted TD to outperform at risk of DLD for story structure and story complexity, following Squires et al. (2014) and against Altman et al. (2016) and Iluz-Cohen & Walters (2011). This prediction might be explained by the specificity of our study, which captures the development of narrative skills at the very beginning. As shown previously, the skills to compose a sequence of cohesive devices develop around the age of seven, when about half of the children are able to produce the complex target GAO episodes, as shown, for example, in Gagarina (2016) for Russian-German bilinguals. Four-year-old children are just beginning to master the organization of narratives. In this early onset TD and at risk of DLD might behave differently in the production of macrostructure components.

As far as the IS – the most language-dependent component of macrostructure – is concerned, we predict significant differences between the groups, based on the consensus in the recent studies on mental language as

associated with Theory of Mind in monolinguals and bilinguals (Miller 2006; Kobayashi, Glover, & Temple 2007) and on the research of lexicon, which was shown to be a vulnerable domain for bilingual DLD (e.g. Kauschke & Klassert, 2014).

4 Method

4.1 Participants

First, the cohort of 167 preschool children was identified on the basis of a short parental query, which included questions about chronological age, age of onset; parental native languages and socio-economic status of the children's family (these children were part of the BIVEM (Berlin Interdisciplinary Alliance for Multilingualism) study). We were looking for children between the age range of two and three, who attended monolingual kindergartens and have at least one native speaker of Russian or Turkish as parent. The children's families should live in different districts of Berlin (children were recruited from 22 full-time monolingual German-speaking Kindergartens) and belong to socio-economic classes ranging from those receiving welfare to the upper middle class.

All children were given the non-verbal IQ test (Snijders-Oomen Nonverbal Intelligence Test (SON-R) 2 ½-7 (Tellegen, Laros, & Petermann, 2007). Parents completed a questionnaire (i.e., *Russian language proficiency test for multilingual children*, Gagarina, Klassert & Topaj, 2010), providing background information about the child. The Kindergarten caretakers filled in a short questionnaire about the conspicuousness of children's (non)verbal behavior. On the basis of these questionnaires and IQ tests, children with hearing, neurological, and cognitive deficits were excluded from the study.

The language skills of the remaining group of children were tested in both languages with a battery of tests (see section 4.2). The children who performed below 1.5 standard deviations on the normed and standardized tests in German (when available, bilingual norms were used) and below 1.25 standard deviations below the mean of the group in their home language (henceforth, L1) – the norm values were not available for these tests – were classified as at risk of DLD (expanding the suggestions by Thordardottir & Brandeker, 2013). In some cases, not all of the different tests per language could be completed; the calculations prior to selection involving these children were based on the remaining data. Supplementing this criterion, the longitudinal observations of two speech-language therapists working with these children were considered. Those children who had less than 10 months of exposure to German were excluded from the study.

By means of these selection criteria, 9 children were identified as being at risk of DLD. Sixteen TD children, matched for age and age of onset, were selected from this bigger cohort as the control group. Table 1 displays the two groups and their mean chronological ages at the time of narrative assessment, age range and age of L2 onset.

Table 1: Means (and Standard Deviations) for chronological age (Age), age range and AoO in months at the time of the LITMUS-MAIN testing

	<u>TD</u> N = 16	<u>at risk of DLD</u> N = 9
Mean Age (SD)	53.4 (7.8)	53.2 (5.6)
Age range	34 – 65	44 – 61
AoO (SD)	9.7 (11.4)	10.4 (10.9)
Max.	28	26
Min.	0	0

Notes. TD = Typical Development; DLD = Developmental Language Disorders; SD = Standard Deviation; AoO = Age of Onset

4.2 Language Tests

All children performed language proficiency tests in both of their languages:

German: For lexicon, production and comprehension of nouns and verbs, and for morpho-syntax, the W-questions subtests from the patholinguistic diagnostics for language disorders (*Patholinguistische Diagnostik bei Sprachentwicklungs-störungen*, PDSS, Kauschke & Siegmüller, 2010) were used. Additionally, the test for checking grammatical understanding (*Test zur Überprüfung des Grammatikverständnisses*, TROG-D, Fox, 2011) of syntax perception was used, along with two different plural tests, one from the *Sprachentwicklungstest für drei- bis fünfjährige Kinder* (SETK 3-5) (Grimm, Aktas & Frevert, 2010) and one from the PDSS (Kauschke & Siegmüller, 2010). Furthermore, two subtests – picture story, W-questions – from the Linguistic language assessment: German as a second language (*Linguistische Sprachstandserhebung – Deutsch als Zweitsprache*, Lise-Daz, Schulz & Tracy, 2011) were administered.

Russian: Russian language proficiency test for multilingual children (SRUK) (Gagarina et al., 2010); Turkish: adaptation of the SRUK (Gagarina et al., 2014 - BIVEM internal material)

The narrative skills were assessed using the LITMUS-MAIN *Multilingual Assessment Instrument for Narratives* (Gagarina et al., 2012; 2015; 2016).

Additionally, the *LITMUS-Sentence Repetition Test* (Hamann et al., 2013) was administered with the group of at risk of DLD and control children.

4.3 Narrative elicitation, transcription and coding

A native speaker of either German or a bilingual investigator administered the LITMUS-MAIN to all bilingual children in a monolingual mode, following the procedure below.

The session starts with a warm-up, comprised of a set of questions like *Tell me about your family, friends, about your favorite game, book or film, etc.*, aimed at getting the child in a talkative mood. The experimenter then places two colored envelopes on the table and asks the child to choose one. From the envelope, the child removes a set of six pictures (connected in a row in a fold-up marionette fashion). The experimenter instructs the child to hand her the pictures, so that she can unfold the pictures in the non-shared attention modus (i.e. the picture sequence is visible only to the child). After the child takes some time to look at the entire picture-story, the experimenter reads the child a story (to accompany the pictures), following up with some questions to the child. After this first story, the experimenter moves to the second part of the assessment. She places a second set of three colored envelopes on the table and asks the child to choose one, saying: *There is a different story in each envelope. Choose one. Now I want you to tell me the story. Look at the pictures and try to tell the best story you can.*

The experimenter helps the child, if necessary, to unfold the pictures. When the story is folded back up by the experimenter so that the child sees only the first two pictures, the child begins telling the story. If the child does not begin on her own, she is prompted as follows: *Tell me the story* (the experimenter points to a picture). When the child has finished ‘telling’ the story in the first two pictures, the experimenter unfolds the next two pictures (so that four pictures are now visible). The experimenter then repeats the process until the child has reached the end of the story. At this point, the experimenter asks ten comprehension questions (for the allowable prompts, see Gagarina et al., 2012:112).

The narratives were transcribed in CHAT format (MacWhinney, 2000) by trained transcribers. A part of the audio files was transcribed in double and inter-rater reliability was checked for. The Cohen’s Kappa reliability score is 90%. Coding and analyses of macro- and microstructure were performed according to the protocols of LITMUS-MAIN, in which the number of possible answers was presented as exhaustively as possible. In this way, inter-reliability for the scoring was guaranteed via the protocols.

4.3.1 Narrative Measures: Macrostructure

For macrostructure, first *story structure* (with a maximal score of 17 points) and *story complexity* (combination of Goals, Attempts and Outcomes, with a maximal score of 6 points per episode indicating the production of a complete GAO episode) were analyzed. The evaluation of *story complexity* was based on Westby's (2005) binary decision tree and included six levels: *no* elements of macrostructure (neither Attempt, A; Outcome, O nor Goal, G) (score of 0); A or O (score of 1); the combination of A and O (score of 2); G (score of 3); GA or GO (score of 4); or the full GAO sequence (score of 6). For this weighted score, the sum of the scores of the three episodes was calculated for each child, resulting in a combined maximal score of 18.

Additionally, all IS tokens and lexemes were counted and the percentage of IS out of all word tokens was calculated. IS included words denoting different internal states: *perceptual state terms* (e.g., see, hear, feel, smell); *physiological state terms* (e.g., thirsty, hungry, tired); *consciousness terms* (e.g. awake, asleep); *emotion terms* (e.g., sad, happy, angry); *mental verbs* (e.g., want, think, know); and *linguistic verbs* (e.g., say, call, ask).

4.3.2 Narrative Measures: Microstructure

For macrostructure, we concentrated on narrative length, measured in the total number of word tokens (WT) without mazes, total number of word types and of verb-based CUs. A CU was defined according to Loban (1976) as “an independent clause with its modifiers”. A CU was considered as verb-based if at least one lexical verb was in it. The sentences with copula, like *The cat is hungry* were not calculated as verb-based CU.

4.4 Sentence Repetition task

The SR task (Haman et al., 2013) consisted of 46 sentences, including the following structures: SV, SVO, sentence brackets, object Wh- questions with Wh-NP and with bare Wh, biclausal relative sentences, coordination with (non-)finite complements, topicalization, subject and object relatives, short and long passives.

For the evaluation of the SR task, we used three types of analyses (based on preliminary findings of Lomako et al., 2015). For all three types, the systematic pronunciation errors were not deemed to be mistakes and were not considered. For the first, more specific analysis, we counted the correct phonemes of each word. If two-thirds of a word were pronounced correctly, the whole word was counted as a correct item (criterion I). This additional evaluation measure was necessary, since some children had problems in the pronunciation and were not able to repeat the long, three- or more-syllable words. For the second, more general analysis, we counted only those words which were completely correct.

For the third analysis, we used a point system, that is, the results of two previous analyses were summed and children received points according to the number of mistakes. For 3 mistakes and more, they got 0 points, for 2 mistakes, 1 point, for 1 mistake, 2 points and for 0 mistakes, 3 points.

5 Results

First, the comparison of the two components of macrostructure, i.e. story structure and story complexity, as well as IS for the two groups, is given and then microstructure is compared. In the second part of the study, the results of SR tests are presented.

The mean scores for the number of story structure elements (maximum 17) – macrostructure analysis – for the two groups are given in Table 2. Between groups, comparisons with the Welch Two Sample t-test showed no significant differences.

For story complexity, measured in weighted scores - a marginally significant difference ($t = -2.117$, $df = 14.108$, $p\text{-value} = 0.052$) was found between the groups (see figure 1). As far as the combination of the episodic components is concerned, at risk of DLD children did not produce either GAO or any GA or GO, whereas more than one third of TD combined Goal with either Attempt or Outcome. No TD children produced the full episodic structure, i.e. GAO.

Table 2: Macrostructure: Story Structure, Structural complexity and Internal States (in tokens, percent and lexemes): Means (and Standard Deviations) for TD and at risk of DLD Group

	<u>TD</u> <i>N</i> = 16	<u>at risk of DLD</u> <i>N</i> = 9	<i>p</i> -Value
Story Structure (<i>SD</i>)	3.5 (1.5)	2.3 (1.7)	0.12
Structural complexity (<i>SD</i>)	3.9 (1.7)	2.1 (2.1)	0.05
Internal States in tokens (<i>SD</i>)	2.6 (2.2)	1.7 (1.1)	0.18
Internal States in percent (<i>SD</i>)	4.0 (3.8)	5.0 (4.0)	0.44
Internal States in lexemes (<i>SD</i>)	1.9 (1.7)	1.6 (1.1)	0.51

Notes. TD = Typical Development; DLD = Developmental Language Disorders; SD = Standard Deviation

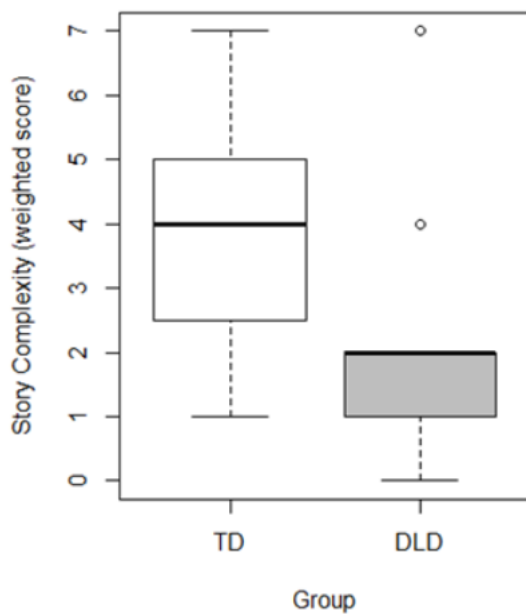


Figure 1: The weighted scores of Story Complexity

Notes. TD Group = typically developing children; DLD Group = children at risk of Developmental Language Disorders.

Story Complexity reflects different levels of a narrative: sequences with no generated Goal statement (A, O, AO); incomplete episodes, which include a Goal statement but lack a complete GAO structure due to omission of an Attempt or Outcome (G, GA, GO), and the most complex and complete episodes including all three GAO components.

The total number of IS, the percent of IS out of all tokens and the number of IS lexemes were slightly higher in TD children, however neither a Wilcoxon rank sum test nor the Welch-two-sample-t-test showed any significant contrast between the groups regarding any IS measure.

An additional correlational analysis between language proficiency and macrostructure was performed: we found positive correlations between the German general language proficiency score and both story structure ($r=0.45$, $t=2.40$, $df=23$, $p=0.02$) and story complexity ($r=0.43$, $t=2.2743$, $df=23$, $p=0.03$). The production of IS measure in tokens and language proficiency in German did not correlate significantly.

For the microstructure, we first compared the number of tokens without mazes (see Table 3). A Welch-two-sample-t-test showed significant differences in story length in TW ($t=-2.24$, $df=21.12$, $p=0.036$), with the TD group telling longer narratives than the at risk of DLD group. The number of different types and CU did not differ significantly across the groups. However, the number of verb-based CU, measured by the Welch Two Sample t-test, was significantly higher in the TD group ($t=-2.17$, $df=22.63$, $p=0.041$), see Figure 2.

Table 3: Microstructure: Length (in Tokens, Types, CU) and Verb-based CU: Means (and Standard Deviations) for TD and at risk of DLD Group

	<u>TD</u> <i>N</i> = 16	<u>at risk of DLD</u> <i>N</i> = 9	<i>p</i> -Value
Total numbers of tokens (<i>SD</i>)	69.8 (34.9)	41.6 (25.0)	0.03
Total numbers of types (<i>SD</i>)	34.3 (11.8)	25.4 (14.1)	0.15
CU (<i>SD</i>)	13.4 (6.7)	9.9 (5.9)	0.21
Verb-based CU (<i>SD</i>)	9.9 (5.3)	6.0 (3.2)	0.04

Notes. TD = Typical Development; DLD = Developmental Language Disorders; SD = Standard Deviation; CU = Communication Unit(s)

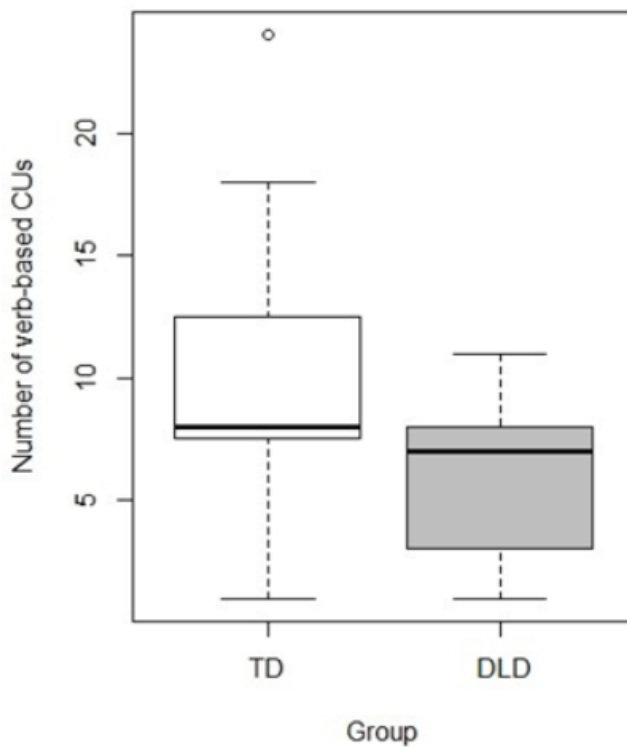


Figure 2: Verb-based Communication Units

Notes. TD Group = typically developing children; DLD Group = children at risk of Developmental Language Disorders.

The Wilcoxon rank sum test was applied to SR task: TD children were significantly better in this task, if two-thirds of a correctly pronounced word was considered to be a correct word ($W=33.5$, $p=0.031$). The analysis of the SR task by whole words was marginally significant ($t=-2.12$, $d=12.39$, $p=0.051$). The in-depth analysis of the most complex structures of SR test, e.g., the sentences with the finite complement clauses, like *Der Prinz will, dass der Ritter die Affen jagt*

‘The prince wants that the knight chases the monkeys’ or *Der Wikinger glaubt, dass die Hexe ihn mag* ‘The viking thinks that the witch likes him’ showed significant outperformance of TD as compared to at risk of DLD children ($t=-2.57$, $df=13.64$, $p=0.02$). See Figure 3.

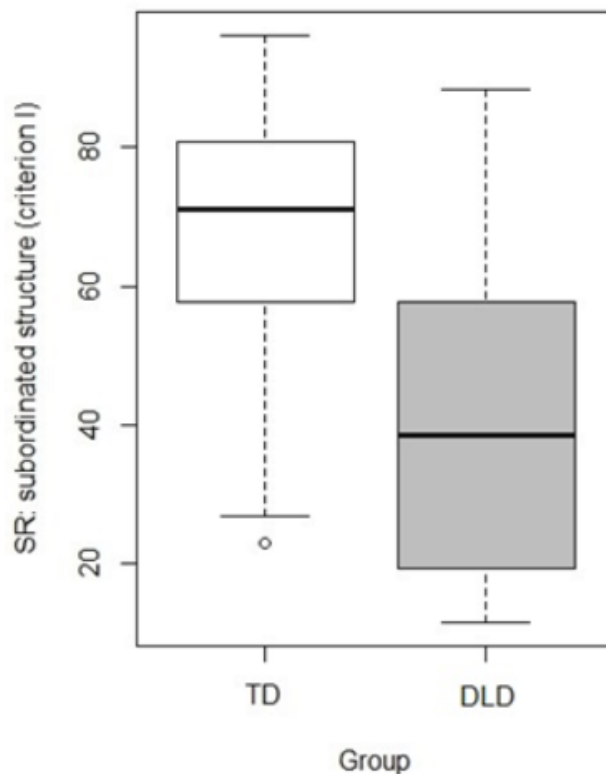


Figure 3: SR: *Sentences with Subordinate Clauses*

Notes. TD Group = typically developing children; DLD Group = children at risk of Developmental Language Disorders.

Sentences containing subordinate clauses are an example of syntactically more complex structures. They are especially relevant in SR tasks as children need to understand the structure in order to be able reproduce the sentence.

6 Discussion and Conclusion

The study compares the elements of macro- and microstructure in the elicited narratives of Russian-/Turkish-German-speaking sixteen TD and nine at risk of DLD bilingual children with a mean age of 4 years and 4 months. The at risk of DLD group was selected out of a bigger cohort of 167 children based on the evaluation of language proficiency in the children’s two languages. For the selection of at risk of DLD we chose two criteria: below 1.5 standard deviations on the standardized tests in German for monolinguals, and 1.25 standard deviations below the mean of the group in the child’s home language. The

narrative and sentence repetition data were collected using the LITMUS-battery of tests, developed within the COST Action IS0806 (Armon-Lotem, Meir & de Jong, 2015).

The TD children showed better performance in all analyzed narrative measures. This outperformance was not, however, always significant. For example, for story structure, no significant differences between TD and at risk of DLD were found. Here our results corroborate recent findings of Iluz-Cohen & Walters (2011) and Tsimpli et al. (2016). Generally, the question whether story structure might cue a distinction between bilingual TD and (at risk of) DLD remains open, because the studies still deliver contradictory results. This might be explained by the fact that, the few bilingual studies dealt with children of various ages and thus different levels of macrostructure and various developmental trajectories across populations. Furthermore, these studies zoom in on a specific age and do not follow the children longitudinally through the development of the macrostructure milestones. Additionally, this might be due to macrostructure itself, which can be said to be of compositional nature rather than monolithic. For example, some elements composing an episode might be cognitively easier to perceive and will be acquired earlier, e.g. an outcome is drawn and explicit and no mental inference is necessary to verbalize it whereas a goal is not drawn and can only be inferred from the pictorial representation of the protagonists' actions. Thus, different components of macrostructure might show various developmental trajectories as far as time and path are concerned. The aggregation of the macrostructure elements of various complexity and perceptibility into one whole might take place at a certain level of episodic complexity when the complex organization of texts is nearly acquired or nearly constituted.

Further, while we have some robust findings on the association of macrostructure across the two languages of bilinguals, irrespective of their proficiency in either language (cf. Gagarina et al., 2015; Pearson, 2002), language-independence of the macrostructure might be weaker at the onset of the narrative skills' development. In order to be able to construct a sequence of cohesive events and verbalize the pictorial content, one needs a critical mass of words. The emergence of early components of macrostructure is therefore dependent on the amount of productive lexical items a child has at her disposal. Our findings on the strong correlation between story structure, story complexity and language proficiency seem to confirm this claim.

Story complexity was marginally significantly lower in at risk of DLD children. Story complexity was suggested to characterize the quality of narratives (Gagarina, 2016) as it measures the level of episodic organization. Since the study investigates the very onset of narrative development, it is remarkable that, already from the onset of the formation of narrative skills, a differentiation is seen. While more than one third of TD children were able to combine the Goal with either

Attempt or Outcome, no at risk of DLD child produced a combination of these two components. So, at risk of DLD children can be said to have less ability in verbalizing a story's universal organization, i.e. a cohesive sequence of events, which is the relatively language-independent domain (Heilmann, Miller & Nockerts, 2010), than their age-matched TD peers.

If macrostructure – at least partially – reflects cognitive abilities, then it should be similar in TD and (at risk of) DLD children. The opposite was found for monolingual children: nine- to eleven-year-old TD children produced more complete episodes than at risk of DLD age-matched peers (Merritt & Liles, 1987). This study, as well as our results, might raise the question of whether cognitive abilities are intact in (at risk of) DLD children, as suggested in the definition of DLD, or whether the domain, responsible for the perception and production of the sequence of cohesive events as represented in a picture stories, is impaired.

Contrary to previous findings on the use of IS in bilingual narratives (Iluz-Cohen & Walters, 2011; Altman et al., 2016 (on retelling); Liles, 1987 for monolinguals), we did not find that the performance of TD was significantly better in at risk of DLD, be it the total number of IS tokens, the percent of IS out of all tokens, or IS lemmas. This could probably be due to the young age of children and the mean length of exposure to German. This was a minimum of ten months, so that the necessary amount of internal state words has not yet been acquired. We might speculate, given the robustness of the other findings, that this difference between TD and (at risk of) DLD will be visible within the course of the narrative skills' development as these children become older.

For microstructure, the traditional measures of story length of in a total number of tokens, total number of types and CU did not reveal any differences, probably due to the children's early age: their narrative skills are at the embryo's level and thus narrative length is basically so short that the differences are not yet significantly visible. However, this variation manifested itself significantly in the number of verb-based CU. This analysis was performed additionally, because we expected to see differences in this domain of the grammatical system. Our expectations followed the studies of Chilla (2011) and Clahsen, Rothweiler, Sterner & Chilla (2014) on the DLD difficulties with verbs in German. Furthermore, verbs constitute the core of a sentence and are triggers for the development of the whole grammatical system (Tomasello, 2003). This finding might give additional evidence in favor of the domain-general interpretation of DLD (Kail, 1994; Miller et al., 2001; Peristeri & Tsimpli, 2013).

Finally, the exact milestones of bilingual story structure development from the emergence of its first components up to the full mastery are not yet defined. This work should be done first in order to be able to identify (at risk of) DLD

macro- and microstructure narrative features (however, for monolinguals see Berman & Slobin, 1994; Trabasso & Nickels, 1992).

To conclude, this study contributes to our knowledge about the onset of the development of productive narrative skills and sheds light on the differences in TD and (at risk of) DLD elicited narratives. It employs a number of criteria to pinpoint bilingual children at risk at the very beginning of their L2 acquisition path. Although the number of the participants in the study is not high, the combination of these selection criteria allows us to derive the conclusion, that it is story length (measured in tokens and verb-based CU) and story complexity that should be included in the composite profile of early, not yet well mastered narrative skills, which differentiate TD and (at risk of) DLD.

7 Acknowledgements

We would like to thank all the children who participated in the study, as well as their parents and teachers for their long-lasting cooperation. The researchers of the Berlin Interdisciplinary Network for Multilingualism, Dorothea Posse and Nathalie Topaj helped us to organize the complex testing procedures while also managing the entire project. Without the help of Kristin Haake, Anne Laporte, Marei Kölling and Levke Schneekloth, the data analyses would not have been possible. No conflict of interests is reported for this study.

8 References

- Altman, C., Armon-Lotem, S., Fischmann, S. & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English–Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, 37(1), 165–193.
- Armon-Lotem, de Jong, J. & Meir, N. (2015). *Assessing Multilingual Children – Disentangling Bilingualism from Language Impairment*. Bristol: UK: Multilingual Matters.
- Berman, R. A. & Slobin, D. I. (Eds.). (1994). *Relating events in narrative: A crosslinguistic developmental study*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bettge, S. & Oberwöhrmann, S. (2012). Migrationshintergrund als Einflussfaktor auf die kindliche Entwicklung im Einschulungsalter – ein multivariates Modell. *Gesundheitswesen* 74(04), 203–209.
- Bishop, D.V.M. & Edmundson, A. (1987). Language impaired 4-year-olds: Transient from persistent impairment. *Journal of Speech and Hearing Disorders*, 52, 156–173.
- Bliss, L.S., McCabe, A. & Miranda, E.A. (1998). Narrative assessment profile: Discourse analysis for school-age children. *Journal of Communication Disorders*, 31, 347–363.
- Chiat, S., Armon-Lotem, S., Marinis, T., Polisenska, K., Roy, P. & Seff-Gabriel, B. (2013). Assessment of language abilities in sequential bilingual children: The potential of

- sentence imitation tasks. In V.C. Mueller Gathercole (Eds.) *Bilinguals and Assessment: State of the Art guide to Issues and Solutions from Around the World* (pp. 56–89). Bristol: Multilingual Matters.
- Chiat, S. & Roy, P. (2013). Early predictors of language and social communication impairments at ages 9-11 years: A follow-up study of early-referred children. *Journal of Speech, Language, and Hearing Research*, 56(6), 1824–1836.
- 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.
- Conti-Ramsden, G., Botting, N. & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (DLD). *Journal of Child Psychology and Psychiatry*, 42(6), 741–748.
- Curenton, S.M. & Justice, L.M. (2004). African American and Caucasian preschoolers' use of decontextualized language: Literate language features in oral narratives. *Language, Speech, and Hearing Services in Schools*, 35, 240–253.
- Fox, A. (2011). *Test zur Überprüfung des Grammatikverständnisses (TROG-D)*. 5. Auflage. Idstein: Schulz-Kirchner Verlag.
- Gagarina, N., Klassert, A. & Topaj, N. (2010). Russian language proficiency test for multilingual children. *ZAS Papers in Linguistics*, 54, Berlin: ZAS.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U. & Walters, J. (2012). MAIN – Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, 56, Berlin: ZAS.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U. & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong & N. Meir (Ed.), *Language Impairment Testing in Multilingual Settings* (pp. 241–274). Amsterdam: John Benjamins.
- Gagarina, N., Klop, D., Tsimpli, I. M. & Walters, J. (2016): Introduction to the Special Issue on Narrative Abilities in Bilingual Children. *Applied Psycholinguistics*, 37(1), 11-17.
- Gagarina, N. (2016). Narratives of Russian-German preschool and primary school bilinguals: rasskaz and Erzählung. *Applied Psycholinguistics*, 37(1), 91-122.
- Grimm, H., Aktas, M. & Frevert, S. (2010). *SETK 3-5. Sprachentwicklungstest für drei- bis fünfjährige Kinder*. Göttingen: Hogrefe.
- Gutiérrez-Clellen, V.F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, 13, 175–197.
- Gutiérrez-Clellen, V.F., Simon-Cerejido, G. & Leone, A. (2009). Codeswitching in bilingual children with specific language impairment. *International Journal of Bilingualism*, 13, 91–109.
- Hamann, C., Chilla, S., Ruigendijk, E. & Abed Ibrahim, L. (2013). A German sentence repetition task: testing bilingual russian-german children. Poster presented at the *COST Action IS0806 Conference*, Crakow.

- Hayward, D. & Schneider, P. (2000). Effectiveness of teaching story grammar knowledge to preschool children with language impairment. An exploratory study. *Child Language Teaching and Therapy*, 16(3), 255–284.
- Heilmann, J., Miller, J.F., 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.
- Iluz-Cohen, P. & Walters, J. (2011). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, 15 (Special Issue 01), 58–74.
- Kail, R. (1994). A method for studying the generalized slowing hypothesis in children with specific language impairment. *Journal of Speech and Hearing Research*, 37, 418–421.
- Kauschke, C. & Klassert, A. (2014). Semantisch-lexikalische Entwicklungsstörungen. In: S. Chilla & S. Haberzettl (Eds.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen* (pp. 171-179) München: Elsevier.
- Kauschke, C. & Siegmüller, J. (2010): *PDSS - Patholinguistische Diagnostik bei Sprachentwicklungsstörungen*. München: Elsevier.
- Kobayashi, C., Glover, G.H. & Temple, E. (2008). Switching language switches mind: Linguistics effects on development neural bases of ‘Theory of Mind’. *Social Cognitive and Affective Neuroscience*, 3(1). 62–70.
- Leonard, L.B. (1998). *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.
- Liles, B.Z. (1987). Episode organization and cohesive conjunctions in narratives of children with and without language disorder. *Journal of Speech and Hearing Research*, 30, 185–196.
- 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, 868–882.
- 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.
- Loban, W. (1976). *Language development: kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.
- Lomako, J., Held, J., Lapenko, I., Gagarina, N. & Lindner, K. (2015). Using sentence repetition for the identification of russian-german bilingual children at risk for SLI. *The 10th International Symposium on Bilingualism (ISB10)*. 20-24 May, 2015. U Rutgers, New Brunswick, NJ, USA.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mandler, G. (1979). Organization and repetition: Organization principles with special reference to rote learning. In L. Nilsson (Eds.), *Perspectives on memory research* (pp. 293–327). Hillsdale, NJ: Erlbaum.

- Marinis, T. & Armon-Lotem, S. (2015). Sentence repetition. In: S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: disentangling bilingualism from Language Impairment* (pp. 95–125). Bristol, Buffalo, Toronto: Multilingual Matters.
- McCabe, A. (1996). Evaluating narrative discourse skills. In: K. Cole, P. Dale & D. Thal (Eds.), *Assessment of communication and language* (pp. 121–141). Baltimore, MD: Paul H. Brooks.
- McCabe, A. & Peterson, C. (1984). What makes a good story? *Journal of Psycholinguistic Research*, 13, 457–480.
- McCabe, A. & Rollins, P.R. (1994). Assessment of preschool narrative skills. *American Journal of Speech-Language Pathology*, 3(1), 45–56.
- Meir, N., Walters, J. & Armon-Lotem, S. (2016). Disentangling SLI and bilingualism using sentence repetition tasks: the impact of L1 and L2 properties. *International Journal of Bilingualism*, 20(4), 421–452.
- Merrit, 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.
- Miller, C. A., Kail, R., Leonard, L. B. & Tomblin, B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 44, 416–433.
- 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.
- Nippold, M.A., Ward-Lonergan, J.M. & Fanning, J.L. (2005). Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic, and pragmatic development. *Language, Speech, and Hearing Services in Schools*, 36(2), 125–138.
- Norris, J.A. & Bruning, R.H. (1988). Cohesion in the narratives of good and poor readers. *Journal of Speech and Hearing Disorders*, 53, 416–424.
- Paradis, J., Genesee, F. & Crago, M.B. (2010). *Dual language development & disorders: A handbook on bilingualism & second language learning*. Baltimore, MD: Paul H.
- Peristeri, E. & Tsimpli, I. M. (2013). Use of reference in the narratives of children with High Functioning Autism and children with Specific Language Impairment. Paper presented at the 4th Pan-Hellenic Conference on Autism, Thessaloniki, Greece.
- Peterson, C. & McCabe, A. (1991). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York, NY: Plenum Press.
- Ringmann, S. & Siegmüller, J. (2013). Die Beziehung zwischen Satzgrammatik und Erzählfähigkeit im unauffälligen und auffälligen Spracherwerb. *Forschung Sprache*, 1, 36–50.
- Clahsen, H., Rothweiler, M., Sterner, F. & Chilla, S. (2014). Linguistic markers of specific language impairment in bilingual children: The case of verb morphology. *Clinical Linguistics & Phonetics*, 28 (9). 1–13.

- Schneider, P., Hayward, D. & Dube, R.V. (2006). Storytelling from pictures using the Edmonton narrative norms instrument. *Journal of Speech Pathology and Audiology*, 30(4), 224–238.
- Schulz, P. & Tracy, R. (2011). *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe.
- Simon-Cereijido, G. & Gutiérrez-Clellen, V.F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexicon and grammar. *Applied Psycholinguistics*, 30, 315–337.
- Skerra, A., Adani, F., & Gagarina, N. (2013). Diskurskohäsive Mittel in Erzählungen als diagnostischer Marker für Sprachentwicklungsstörungen. In T. Fritzsche, C.B. Meyer, A. Adelt & J. Roß (Eds.), *Spektrum Patholinguistik* 6, 27–158.
- Squires, K.E., Lugo-Neris, M.J., Peña, E.D., Bedore, L.M., Bohman, T.M. & Gillam, R.B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language and Communication Disorders*, 49, 60–74.
- Stein, N.L. & Glenn, C.G. (1979). An analysis of story comprehension in elementary school children. In R.O. Freedle (Eds.), *Discourse processing: Multidisciplinary perspectives*. Norwood, NJ: Ablex.
- Swanson, L., Mills, C., Hood, L. & Fey, M. (2005). Use of narrative-based language intervention with children who have specific language impairment. *American Journal of Speech-Language Pathology*, 14, 131–143.
- Tellegen, P.J., Laros, J.A. & Petermann, F. (2007). *Snijders-Oomen Non-verbaler Intelligenztest von 2½ bis 7 Jahren (SON-R 2½–7)*. Göttingen: Hogrefe.
- Thordardottir, E. & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, 46, 1–16.
- Torrance, N. & Olson, D.R. (1984). Oral language competence and the acquisition of literacy. In A. Pellegrini & T. Yawkey (Eds.), *The development of oral and written language in social contexts*. Nordwood, NJ: Ablex.
- Trabasso, T. & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, 15, 249–275.
- Tsimpli, I. M., Peristeri, E. & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, 37(1), 195-216.
- Wallach, G.P. (2008). *Language intervention for school-age students: Setting goals for academic success*. St. Louis, MO: Mosby.
- Westby, C.E. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and reading disabilities* (pp. 157–232). Boston, MA: Allyn & Bacon.
- International Organization for Migration (Eds.) (2013). WMO report - World Migration Report 2013.