

**Narrative Generation and Narrative Recall Recruit Different Executive Functions
in Preschoolers with and without Developmental Language Disorder**

Yuanyuan Lin^a, Li Sheng^b, Huanhuan Shi^c, Wenjie Yan^a, Yiwen Zhang^{a*}

*^aDepartment of Developmental and Behavioral Pediatrics, Shanghai Children's
Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China;*

*^bResearch Centre for Language, Cognition, and Neuroscience & Department of
Chinese and Bilingual Studies of the Hong Kong Polytechnic University, Hong Kong,
China; ^cNew York University, New York, United States*

*corresponding author: Yiwen Zhang, zhangyiwen@scmc.com.cn

Yuanyuan Lin and Li Sheng contributed equally as co-first authors.

Narrative Generation and Narrative Recall Recruit Different Executive Functions in Preschoolers with and without Developmental Language Disorder

Children with developmental language disorder (DLD) are poor at story-telling and show weaknesses in various executive functions (EFs). Narrative tasks are frequently used in clinical assessment to capture the linguistic vulnerabilities of individuals with DLD. But we know little about the demands of different narrative tasks on EFs. This study explores the relationship between EFs in a daily life context and performance on two narrative tasks. Fourteen Mandarin-speaking preschoolers with DLD and 34 typically-developing (TD) controls completed a story generation and a story recall task. Their parents filled out the Behavior Rating Inventory of Executive Functions-Preschool (BRIEF-P). The TD group outperformed the DLD group on narrative macrostructure and microstructure, and the inhibit, shift, and global executive composites of the BRIEF-P. On the story recall task, after controlling for standardized language test scores, working memory scores explained unique variance in both narrative macrostructure and microstructure performance. On the story generation task, after controlling for language skills, macrostructure performance was predicted by inhibit, working memory, and plan/organization composites, and microstructure performance was predicted by the inhibit composite. Narrative recall relies heavily on working memory capacity as children must recall the details provided in the mature adult model; narrative generation requires multiple EFs as children must plan the organization of story elements, selectively attend to relevant visual details in the pictorial stimuli, and monitor their own language production. The findings have implications for understanding the sources of language difficulties in DLD and the selection of narrative task in clinical practice.

Keywords: developmental language disorder; executive functions; narrative; preschoolers

Introduction

Developmental language disorder (DLD) is a neurodevelopmental disorder characterized by poor language skills not explained by intellectual disability, sensory, and overt neurological impairment (Bishop et al., 2017). Specific Language Impairment (SLI) is another term frequently used to describe children who exhibit significant limitations in language learning despite having typical intelligence, hearing, and no diagnosed neurological or developmental disorders. While DLD and SLI overlap a great deal, there are fine distinctions in the diagnostic criteria and the theoretical perspectives on the nature of the language difficulties. The DLD label stipulates a nonverbal intelligent quotient (IQ) cutoff of 70, whereas the SLI label often requires a higher IQ cutoff (e.g., 85) to reflect a “purer” language impairment. DLD is also perceived as a broader label as it explicitly encourages the study of children who have common comorbid disorders (e.g., developmental dyslexia, development coordination disorder). The linguistic deficits of children with DLD are manifested across multiple levels, including vocabulary, grammar, verbal learning, and connected discourse. Moreover, many children with DLD demonstrate subclinical deficits in various executive functions (e.g., Vissers et al., 2015). The current study explores the relationship between executive functions in a daily life context and functional expressive language (i.e., narrative) skills among Mandarin-speaking children with DLD and their typically developing (TD) peers.

Narrative language in children with DLD

Narrative is an authentic, ecologically valid, and versatile form of assessment (Justice et al., 2010). Narrative samples generate rich information about the speaker’s ability to integrate vocabulary, grammar, and pragmatic rules all at once. There is a rapidly

growing literature on narrative skills in various clinical populations, including but not limited to children with DLD (e.g., Charest et al., 2020), Autism Spectrum Disorders (e.g., Durrleman et al., 2022), and hearing impairment (e.g., Walker et al., 2023). Children with DLD present weaknesses in multiple dimensions of narrative ability including macrostructure (i.e., the content of the story, also called story grammar elements) and microstructure (i.e., the form of the story, or the use of linguistic devices). Several systematic reviews appeared in recent years summarizing narrative performance patterns in monolingual and bilingual children with DLD (Andreou & Lemoni, 2020; Ebert, 2020; Srivastava et al., 2023; Winters et al., 2022) attesting to the popularity of this language assessment technique. These reviews noted many sources of variations in this literature, including the story materials, the elicitation method, the selection of macrostructure and microstructure narrative measures, as well as the age, language, and profiles of the participants. Nevertheless, across studies and languages, there are convergent patterns. For instance, Winters et al. (2022) reviewed 37 studies across languages and found that the greatest differences between the DLD and TD groups were in story macrostructure and grammatical accuracy within microstructure.

In Mandarin-speaking children with DLD, the focal population of the current study, a handful of narrative studies exist (e.g., Hao et al., 2018; Lai & Wang, 2017; Sheng et al., 2020). A recent systematic review of DLD in Chinese children (Sheng et al., 2023) indicates that across studies, a common finding was that the DLD group scored significantly lower than TD peers on story macrostructure. With regard to narrative microstructure, group differences are more consistently found in measures of lexical diversity (i.e., number of different words) and sentence length (i.e., mean length of utterances) but less consistently found in measures of productivity (e.g., total number of words) and grammatical accuracy.

There are many ways to elicit narratives from young children (Boudreau, 2008). One of the most common methods is through the use of wordless picture books, e.g., the Frog Stories (Berman & Slobin, 1994) and the sequenced pictures from the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2019). This method is advantageous when working with young children because the pictorial support keeps the memory load at a manageable level and the sequence in the pictures provides organizational guidance in the narration. Both story recall and story generation are frequently elicited from young children. Story recall requires that the child first listens to an adult model and then recall the story; story generation requires the child to generate their own stories. Both TD children and children with DLD tend to benefit from the adult model and produce longer, more complete, and linguistically more sophisticated stories in the recall condition than the generation condition (Kunnari et al., 2016; Otwinowska et al., 2020; Roch et al., 2016; Sheng et al., 2020). However, in studies that directly compared a TD group and a clinical group (Merritt & Liles, 1989; Sheng et al., 2020; Walker et al., 2023; Westerveld & Gillon, 2010) on both narrative recall and generation tasks, findings have been mixed regarding whether recall or generation shows greater TD-clinical group difference. These differences are difficult to reconcile because the studies differed in terms of participant profile (e.g., DLD, reading disability, hearing impairment), stimuli (e.g., length of the story and complexity of the plotline) and procedures (e.g., instruction, task order). But these variations invite further investigations into the underlying cognitive demands of different narrative tasks to deepen our understanding of the cognitive factors that may contribute to functional communication performance in various disordered groups.

Executive functions in children with DLD

Executive functions (EFs) encompass a broad range of higher-order cognitive processes that play a critical role in our everyday lives (Diamond, 2013). It is generally accepted that EFs are made up of three core components: inhibition, working memory (WM), and cognitive flexibility (also called shifting) (Miyake et al., 2000). Inhibition is the ability to control one's attention, avoid distraction, and inappropriate verbal and nonverbal responses. WM refers to one's ability to hold information in memory and simultaneously perform mental operations on this information. Shifting refers to switching back and forth between different tasks and operations. Beyond these, planning is considered a higher level EF which is based on these three core elements. Planning is the ability to set a goal, foresee the steps involved, and prioritize the steps towards achieving the goal (Diamond, 2013; Miyake et al., 2000).

Although the hallmark deficit of DLD is poor language, there is considerable evidence that weaknesses in EFs are also part of the DLD profile (Kapa et al., 2017; Kapa & Plante, 2015; Pauls & Archibald, 2016; Vissers et al., 2015). Specifically, in the domain of inhibitory control, researchers often employ experimental paradigms such as the go/no go task, the stop-signal Task, the flanker task, and the stroop task, among others, to assess children. Studies suggested that children with DLD generally performed worse than TD peers on these tasks (Kapa et al., 2017; Roello et al., 2015; Spaulding, 2010), although null findings were also noted (e.g., Peristeri et al., 2024¹).

¹ Peristeri et al. (2024) included six groups: TD monolinguals, TD bilinguals, ASD monolinguals, ASD bilinguals, DLD monolinguals, and DLD bilinguals. Their focus is on bilingualism; therefore, a direct contrast was not provided between the TD monolingual and

For Working memory, many studies have identified that children with DLD performed worse compared to their TD peers (Henry & Botting, 2017; Vugs et al., 2014, 2016, 2017), with verbal working memory deficits being two to three times greater than visual-spatial working memory deficits (Vugs et al., 2013). Research on shifting tasks in preschool children with DLD is limited. Yang and Gray (2017) found that children with DLD significantly underperformed in verbal shifting tasks compared to TD children, a finding consistent with Farrant et al. (2012). However, in non-verbal shifting tasks, there was no difference in performance between children with DLD and TD children. It is important to note that 45.5% of children with DLD were excluded from the study due to their inability to complete the pre-shift tasks, compared to only 6.7% of TD children. This suggests that children with DLD also face difficulties in completing non-verbal shifting tasks (Yang & Gray, 2017). To sum up, DLD seems to negatively affect various components of EFs across linguistic and nonlinguistic domains and auditory and visual-spatial modalities.

Besides standardized neuropsychological assessments and experimental tasks, EFs can be measured using standardized rating scales to collect information from parents and teachers regarding children's behaviors in daily life context. Wittke et al. (2013) administered the Behavior Rating Inventory of Executive Function-Preschool Version (Sherman & Brooks, 2010) to parents and teachers of 38 three- to five-year old children with DLD and TD. The DLD group scored significantly higher (i.e., indicating weaker skills) than the TD group on the emergent metacognition index (comprising scores on the WM and plan/organize scales) and the flexibility index (comprising scores

DLD monolingual groups. However, Table 3 (p.217) suggests that these two groups were not different in the efficiency of inhibition on the EF task.

on the emotional control and shift scales). Moreover, the global executive composite scores on the BRIEF-P significantly correlated with children's standardized language test scores. As preschoolers may have difficulties completing some of the standardized EF tasks due to attentional and motor constraints, behavioral rating scales are especially useful assessment tools to help us understand whether the EF limitations documented among young children with DLD through decontextualized tasks are also noticeable in their daily living.

Executive functions and narrative language performance

More recently, researchers have begun to examine the contribution of linguistic and domain-general cognitive skills to narrative production in children with DLD, ASD, and hearing impairment (e.g., Dodwell & Bavin, 2008; Durrleman et al., 2022; Smolak et al., 2020; Walker et al., 2023). Because our primary focus is on DLD and on studies that included executive functions and different types of narrative tasks, below we expounded on the studies that reported the most relevant findings.

Blom and Boerma (2016) conducted a longitudinal study with Dutch-speaking children (who were 4 to 7 years old) with and without DLD. At time 1, children were assessed on structure language (vocabulary, grammar), narrative (story generation macrostructure), sustained selective attention (SSA), and two memory tasks. SSA refers to the ability to maintain one's attention to a specific task across time and is believed to be a more basic EF skill that underlies all EF abilities (Garon et al., 2008; but see Kapa et al., 2017 for different findings). At time 2, children were re-assessed on the story generation task. Of all the time 1 language and cognitive measures, only SSA was a significant predictor of time 2 narrative macrostructure score. Moreover, mediation

analysis suggested that SSA mediated the relationship between group status and story macrostructure production.

While Blom and Boerma (2016) documented a specific link between attention (but not memory) and story generation performance, Duinmeijer et al. (2012) was the first to examine how cognitive skills relate to different narrative tasks. They administered story recall (*The Bus Story*), story generation (*The Frog Story*), sustained auditory attention (SAA), and verbal memory (VM) tasks to Dutch-speaking children with and without DLD ages 6 to 9 years old. Correlational analyses with story macrostructure scores showed distinct patterns between the two narrative tasks: on the story generation task, macrostructure was related to SAA, but not VM; on the story recall task, macrostructure was related to VM but not SAA.

Taken together, these two studies suggest that story recall relies heavily on memory capacity as children engage in recalling the details provided in the mature adult model (Duinmeijer et al., 2012); story generation is both concurrently and longitudinally related to continuous focused attention as children have to stay on task and selectively attend to relevant visual details in the pictorial stimuli (Blom & Boerma, 2016; Duinmeijer et al., 2012).

The current study

Previous studies yield robust evidence of deficits on narrative macrostructure and microstructure performance in children with DLD (Andreou & LEMONI, 2020; Ebert, 2020; Sheng et al., 2020, 2023; Winters et al., 2022). These children also show weaknesses in various EF components (e.g., Kapa et al., 2017; Vissers et al., 2015; Vugs et al., 2013; Wittke et al., 2013). Moreover, emerging evidence suggests that different narrative tasks may require different EF skills (Blom & Boerma, 2016;

Duinmeijer et al., 2012). To further our understanding of the complex relationships among various EF skills and narrative language, we administered measures of EF and two narrative tasks to preschoolers with DLD. There are two main methodological differences between our study and previous studies. First, we used tightly controlled and equated stories from the MAIN (Gagarina et al., 2019) for the narrative recall and generation tasks such that differences in the involvement of EF skills cannot be attributed to differences in task materials. Second, while previous studies of EF-narrative relationships used decontextualized experimental tasks, we measured EF skills using the BRIEF-P, which evaluates multiple EF components in daily life context. These methodological modifications allow us to see if the previously documented relationships are generalizable to new contexts and new materials.

The research questions are: 1) Do preschoolers with DLD show poorer performance on narrative macrostructure and microstructure than TD peers? 2) Do preschoolers with DLD show weaker executive functions in daily life context than TD peers? 3) What is the relationship between children's EF skills and their narrative recall and narrative generation performance? Based on existing evidence, we hypothesized that: 1) children with DLD would perform more poorly than TD peers on both narrative macrostructure and microstructure (Andreou & LEMONI, 2020; Ebert, 2020; Sheng et al., 2020, 2023; Winters et al., 2022); 2) children with DLD would show generally weaker performance on the EF components of the BRIEF-P (Kapa et al., 2017; Vissers et al., 2015; Vugs et al., 2013; Wittke et al., 2013). However, given the limited evidence on DLD using rating scales, we did not have a priori hypotheses regarding which specific components would be significantly weaker; and 3) the memory component of the EF would be related to narrative recall, whereas the attention component would be related to narrative generation (Blom & Boerma, 2016; Duinmeijer et al., 2012).

Method

Participants

Fourteen children with DLD and 34 TD children participated in the study. The two groups were comparable in mean age, maternal education level, and amount of Mandarin exposure but differed significantly on standardized language test scores and performance IQ. Participant information is presented in Table 1. This research was approved by the institutional review board (IRB) of the first author's affiliation (approval number: SCMCIRB-K2017016).

Children with DLD were recruited from the Department of Developmental and Behavioral Pediatrics of a major pediatric hospital in China. Children were brought in by their caregivers to the outpatient clinic. If the caregiver expressed concerns about the child's language skills, the child-caregiver duo would be interviewed by an experienced developmental and behavioral pediatrician and the child would undergo language and cognitive testing. To be included in the DLD group, the child had to meet the following criteria: (1) have at least one composite standard score lower than 80 on the Diagnostic Receptive and Expressive Assessment of Mandarin-Comprehensive (Ning et al., 2014). This criterion was adopted based on evidence of optimal diagnostic accuracy published by the test developers (Liu et al., 2017); (2) have a performance intelligence quotient (IQ) of 85 or higher on the Wechsler Preschool and Primary Scale of Intelligence-Revised (Wechsler, 2018); (3) passed a hearing screening; and (4) did not have a diagnosis of neurological disorder, genetic disorder, cerebral palsy, or autism spectrum disorder.

Among the 34 TD children, five were recruited from the hospital when these children were brought in by their parents to the outpatient clinic for non-language related concerns (e.g., physical growth, nutrition). Because of the difficulties in finding

TD children in a hospital setting, the remaining 29 TD children were recruited via social media (i.e., WeChat) advertisements. To be included in the TD group, all five composite scores on the DREAM-C had to be ≥ 80 and performance IQ on the WPPSI-R had to be ≥ 85 .

The primary caregivers of all participants signed an informed consent and completed a questionnaire regarding the child general developmental history, parental education level, and children's exposure to Mandarin.

Materials

BRIEF-P

The BRIEF-P is a questionnaire filled out by parents to evaluate executive functions in children between the ages of 2 years and 5 years 11 months. The questions focus on individual's behaviors in everyday environment (Sherman & Brooks, 2010). A parent-report method to assess EFs of preschool-aged children was chosen for two reasons. First, for children with DLD, completing laboratory-based EF tasks can be challenging because many EF tasks rely on verbal instructions and responses, which may pose a significant barrier for children with impaired language comprehension and production. Consequently, their performance on EF tasks may not solely reflect their executive capabilities but also the limitations imposed by their language disorders. Second, parents can provide a comprehensive perspective on their children's behaviours across different settings (e.g., home, school, peer interactions), which offers richer information than behaviours observed in a single lab setting. For example, laboratory settings are often quiet and almost distraction-free. To encourage children to complete tests, experimenters might provide encouragement or reminders when children's attention wanes, which can act as a form of "prefrontal" supervision.

The BRIEF-P includes 63 items rated on a Likert scale of one to three (never, sometimes, often). The items contribute to five subscales: inhibit, shift, emotional control, working memory, and plan/organize. The global executive composite (GEC) is derived from the above five subscales to index children's overall executive functions. Higher scores indicate worse performance. The original BRIEF-P is in English, and it has been adapted into different languages, including Chinese, to suit various populations globally. The process of translating the BRIEF-P into Chinese goes beyond simple language translation; it involves cultural adaptation to ensure that the items are culturally relevant and understandable. It was translated by two PhD students and then proofread by two individuals proficient in English. Subsequently, it was back-translated by people majoring in English who had not previously seen the questionnaire, and finally, it was sent to the original authors for their final approval, ensuring the quality of the translation. The BRIEF-P has adequate internal reliability and adequate content validity and construct validity in Chinese children, e.g., the internal consistency reliabilities ranged from 0.78 (plan/organize) to 0.90 (inhibit) (Lu et al., 2017).

DREAM-C

The DREAM-C is a standardized language assessment for Mandarin-speaking children ages 2;6 to 7;11. The test was administered by professionally trained personnel using an iPad and standard scores were automatically generated by the iPad App. The test is organized into seven subscales (vocabulary, fast mapping, closed class, logical expressions, sentence contrasts, event description and sentence repetition) that are used to estimate five composite scores (total language, receptive language, expressive language, semantics and syntax) (Liu et al., 2017).

MAIN

The MAIN (Gagarina et al., 2019; Luo et al., 2020) is designed to assess narrative skills for children aged 3 to 10 years. The Mandarin version was adapted by native speakers of Mandarin in collaboration with the original author of the MAIN, following specific guidelines for adapting MAIN into new languages (Luo et al., 2020). The instrument includes four stories, which are parallel in terms of cognitive and linguistic complexity, matching of macro- and micro- structures, cultural appropriateness and test robustness. Each wordless story contains six full-colour pictures. The MAIN manual provides detailed instructions on task administration and coding procedure.

WPPSI-R

The WPPSI-R is a standardized normative measure of cognitive ability for 4 to 7-year-olds. The Chinese version of the WPPSI-R has shown evidence of good reliability (Zhu YM, et al., 1984). The WPPSI-R includes a verbal IQ subscale and a performance IQ (PIQ) subscale. The PIQ subscale evaluates non-verbal and spatial reasoning abilities through five subtests: animal pegs, picture completion, mazes, geometric design, and block design. In the current study, we used all five subtests from the PIQ subscale as these tests primarily assess non-verbal skills and minimize reliance on language abilities. These scores were scaled for age (mean \pm SD of 100 ± 15).

Procedures

Children was tested in a quiet assessment room. Tasks were administered in the following order: hearing screening, DREAM-C, WPPSI-R, and MAIN. The story generation task (the Bird or Goat stories) was always given before the story recall task (the Cat or Dog stories). According to the manual of MAIN (Gagarina et al., 2019; Luo

et al., 2020), we implemented a sorting procedure to counterbalance the potential bias caused by narratives, wherein children were assigned to a predetermined order based on their entry into the study (i.e., the first participant received the Bird story for generation task and the Cat story for recall task, the second participant received the Goat story for generation task and the Cat story for recall task, the third participant received the Bird story for generation task and the Dog story for recall task, the fourth participant received the Goat story for generation task and the Dog story for recall task, and so forth). In the story generation task, they had to tell their own story based on the pictures. In the story recall task, the experimenter read the story script aloud while children were reviewing the pictures. After that, children were required to retell the story with the aid of the pictures. Children's narrations were audio-recorded.

Transcription and coding

Using the Computerized Language Analysis (CLAN) program, all samples were transcribed verbatim into Chinese characters in Codes for Human Analysis of Transcripts (CHAT) by the first author (MacWhinney, 2018). Utterances were segmented into clauses according to the guidelines in Sheng et al. (2020). Word segmentation was performed using the *Chinese online word segmentation system* (<https://ckip.iis.sinica.edu.tw>) and then checked manually.

The scoring sheet of the MAIN consists of three parts: story structure score, structural complexity, and internal state terms (IST, which include words that describe perceptual state e.g., hear; physiological state, e.g., thirsty; consciousness, e.g., awake; emotion, e.g., happy; mental verbs, e.g., think; and linguistic verbs, e.g., tell). The story structure score ranges from 0 to 17 and consists of the story setting (time and place), and 3 episodes (i.e., IST as initiating event, goal, attempt, outcome, and IST as

reaction). Structural complexity evaluates the completeness of each episode and can be classified into the following: goal only (G); attempt and outcome (AO); goal and attempt/goal and outcome (GA/GO); goal, attempt and outcome (GAO). For structural complexity, the maximum score is 3 as there are only 3 episodes. The IST score equals the number of ISTs in the child's narrative. In addition to the above macrostructure elements, we calculated three microstructure scores including mean length of utterance (MLU), total number of words (TNW), and number of different words (NDW) via the CLAN program.

Reliability

Twenty percent of the samples were independently transcribed, coded and scored by the third author. Reliabilities were calculated based on the following formula:

$(M1+M2)/2*100$ (M1=number of consistencies/number of the first transcriber; M2=number of consistencies/number of the second transcriber). The reliability was 94.31% for narrative transcription, 96.67% for utterance segmentation, 92.21% for word segmentation, and 93.26% for MAIN scoring. All inconsistencies were resolved by discussion.

Data analysis

Data analysis was completed in R (version 4.2.1). To test normality, all data sets were subjected to the Shapiro-Wilk test. Data with normative distributions were expressed as means and standard deviations, while data with non-normal distributions were presented in quartiles. Continuous variables that follow a normal distribution were subjected to comparison using the Student *t* test. Effect sizes were estimated using Hedges' *g*. Non-normally distributed variables were compared using the Mann-Whitney *U* test. Effect sizes were estimated using rank-biserial correlation (*r*). For categorical variables, the

Chi-square test was used and effect sizes were estimated by Cramer's V .

A power analysis using the *pwr* package indicated that current sample size would yield 70% power to detect an effect estimated at medium to large with a one-tailed alpha of 0.05. In addition, we also applied a multiple regression power calculation, using an effect size (f^2) ($f^2=R^2/(1-R^2)$), sample size of 48 (total), and an alpha error of 0.05. Results showed that our study had more than 89% power to demonstrate the observed relationship.

For the first research question, a robust two-way (group x task) mixed ANOVA based on trimmed means, specifically designed for non-normally distributed data, was performed with the *bwtrim* function in WRS2 package (Mair & Wilcox, 2020). For the second research question, Student *t* or Mann-Whitney U tests were used to compare between the groups. The third research question was addressed by employing linear regression through the *lm* function in the stats package. Before conducting linear regression, we utilized the *gvlna* function to comprehensively check the extent to which the linear regression model satisfied basic assumptions, including linearity, homoscedasticity, independence of error terms, and normal distribution of error terms. All but one analyses conformed to the assumptions of linear regression. Only when the independent variable was the total number of words, we found that the assumption of the normal distribution of error terms was not met (p between 0.01 and 0.04). Upon further verification using QQ plots, the data distribution was almost aligned in a straight line. Therefore, we kept the results of the linear regression analysis. This analysis aimed to explore the relationship between narrative production and EF while controlling for structural language scores on the DREAM-C. Because of the significant interrelations between EF and structural language (e.g., Wittke et al., 2013), and between structural language and narrative skills (Fiani et al., 2022), we deemed it necessary to control for

variations in structural languages skills to better isolate the effect of EF on narration. According to the manual of the DREAM-C test, its semantics subscale encompasses vocabulary, fast mapping, closed class and logical expressions while the additional subscales cover sentence contrasts and event descriptions, elements regarded as advanced language skills (Liu et al., 2017). Previous studies focusing on understanding the contributions of lower-level language skills to narrative skills found that vocabulary was a significant contributor to narrative skills (Bitetti et al., 2020; Khan et al., 2021). Apart from the above, given the exploratory nature of this question, we chose the semantics composite of the DREAM-C to index structural language because it has the smallest effect size of the five composites (see Table 1) and could potentially maximize our chance of finding an effect. $P < 0.05$ was accepted as statistically different.

Results

Narrative production

Macrostructure

Table 2 shows descriptive data on narrative macrostructure performance. The TD group earned higher scores than the DLD group on all three measures of macrostructure. The recall task elicited numerically higher scores than the generation task in all but one instances (TD group's structural complexity score).

For the story structure score, we detected a main effect of groups ($Q=20.59$, $p<0.001$), with the TD group outperforming the DLD group; a main effect of task ($Q=29.26$, $p<0.001$), indicating better performance on the story recall task than the generation task; and a non-significant interaction between group and task ($Q=1.05$, $p=0.32$). For the structural complexity score, we detected a main effect of group ($Q=7.34$, $p=0.02$). There was no main effect of task ($Q=1.84$, $p=0.20$), nor was there an

interaction between group and task ($Q=0.88, p=0.36$). For the IST score, we detected a main effect of group ($Q=10.39, p=0.005$), a main effect of narrative task ($Q=14.88, p=0.001$), and a significant interaction between group and task ($Q=6.08, p=0.03$). The TD group outperformed the DLD group regardless of narrative task ($ps=0.002$). The TD group showed better performance on narrative recall than the generation task ($p<0.001$), but no difference was observed between tasks in the DLD group.

Microstructure

Descriptive data are displayed in Table 3. Overall, the TD group outperformed the DLD group on all three measures, and the recall task elicited longer utterances and greater number and diversity of words than the generation task. For MLU, there was a main effect of group ($Q=17.69, p<0.001$), while the effect of task ($Q=3.96, p=0.06$) and the interaction between group and task were not significant ($Q=0.11, p=0.74$). For TNW, the main effects of group ($Q=10.39, p=0.005$) and task were significant ($Q=16.15, p=0.001$), while the interaction was not ($Q=4.24, p=0.06$). For NDW, the main effects of group ($Q=15.26, p=0.001$) and task ($Q=19.10, p<0.001$) were significant, while interaction between group and task was not ($Q=3.24, p=0.09$).

Executive functions

Table 4 presents the two groups' scores on different components of the BRIEF-P. The DLD group showed higher scores (i.e., weaker performance) than the TD group on all components. The group difference reached statistical significance for inhibit, shift, and GEC, with small to medium effect sizes. The group difference approached significance for the working memory component with a medium effect size.

Relationship between narrative production and executive functions

Linear regressions were performed with the story structure score, MLU, TNW, and NDW as the predicted variables and the five EF components as predictors for the two narrative tasks, respectively. The two groups were pooled together given the small sample size. Results are shown in Table 5. For story generation, after controlling for structural language scores derived from the DREAM-C, Inhibit, WM, and Plan/Organize predicted story structure score; Inhibit also predicted all three microstructure scores (i.e., MLU, NDW, TNW). For story recall, WM was the only significant predictor and explained unique variance in macrostructure (story structure score) and the three microstructure scores.

Discussion

We compared language and EF performance between DLD and TD groups using functional communication tasks (narratives) and contextualized measures of EF (parent report). We also explored the relationship between the two sets of language and EF measures. In the following sections, we discuss the findings in relation to the three research questions.

Narrative performance

Consistent with our hypothesis, we found a main effect of group on all measures of narrative macrostructure and microstructure. The group effect was significant even for measures that previously failed to show DLD-TD differences such as the story complexity score (e.g., Altman et al., 2016; Sheng et al., 2020; Tsimpli et al., 2016) and TNW (e.g., Hao et al., 2018; Sheng et al., 2020). When effect sizes were examined, among the macrostructure measures (Table 2), the story structure score of the MAIN showed the largest and most stable group differences, whereas the IST and story

complexity scores were less differentiating. Among microstructure measures (Table 3), the productivity measure of TNW was less differentiating than the sentence length (MLU) and lexical diversity (NDW) measures. These patterns are generally consistent with findings reported in Sheng et al.'s (2023) systematic review of Chinese DLD.

Though task effect was not a main focus of the current study, a significant task effect was identified in story structure score, IST, TNW, and NDW. In all cases, better performance was achieved for story recall than story generation. This is consistent with previous studies that used the MAIN task materials and compared the effect of different narrative elicitation method in TD children (e.g., Kunnari et al., 2016; Otwinowska et al., 2020; Roch et al., 2016) as well as studies that examined task effect in various clinical groups (e.g., Merritt & Liles, 1989; Sheng et al., 2020; Walker et al., 2023; Westerveld & Gillon, 2010). Note that we used a fixed generation-then-recall task order, while in Sheng et al., (2020), a fixed recall-then-generation task order was used. Both the current study and the Sheng et al. study used the MAIN stories. Thus, the recall advantage in the current study could not be simply attributed to a practice effect but was more likely due to the availability of the story model, which is believed to provide a more supportive context for all children

Executive functions

As hypothesized, children with DLD had worse performance on EF components than TD peers. More specifically, we found significant group differences on the inhibit, shift, and global executive composite dimensions of the BRIEF-P. Difference in working memory approached significance. In comparison to TD peers, children with DLD in the current sample were perceived by their parents to be more forgetful, and poorer at suppressing impulses, resisting interferences, stopping undesirable behaviours, shifting from one activity to the next, remembering instructions and completing multi-

step activities. Previous research has consistently shown that children with DLD exhibit executive functioning (EF) challenges in daily life. In a previous study with English-speaking preschoolers, Wittke et al. (2013) found poorer performance in the DLD group than the TD group on the emergent metacognition index (WM & plan/organize scales) and the flexibility index (emotional control & shift scales). Trainor (2012) also found that children with DLD had poorer performance on the BRIEF-P scales for inhibit, plan/organize, shift, and working memory in a United States preschool population. Genenbacher's study (2013) extends these findings by utilizing teacher evaluations instead of parental ratings. Significant differences were observed between the executive functions of the typical language group and those of the language-impaired and at-risk groups, particularly in the domains of inhibit, working memory, and plan/organize. These consistent results across different evaluators underscore the broader impacts of DLD on cognitive processes beyond language skills. The current findings add to the larger literature documenting domain-general cognitive deficits in the DLD group and were among the first to report EF impairments in Mandarin-speaking children with DLD.

Why might children with DLD, whose primary deficiencies lie in the language domain, show deficits in various EFs? While we are unable to draw any causal relationship between impaired language functions and EFs, extant theoretical models have posited several potential pathways between language and EFs (Kapa & Mettler, 2021). The unidirectional model with EF influencing language has received some support: a predictive relationship between early EF scores and later language abilities (e.g., vocabulary, grammar) was documented in TD children (Verhagen & Leseman, 2016; Weiland et al., 2014). Moreover, intervention focused on WM training resulted in improved WM as well as improved listening comprehension in preschoolers from low

SES backgrounds (Rojas-Barahona et al., 2015). The opposite unidirectional model should also be considered. Vallotton and Ayoub (2011) found that a child's expressive vocabulary at age 2 predicts later growth in EF which suggests that language abilities may drive the development of EFs. Language is often used as a tool to facilitate thinking, problem-solving, and self-regulatory processes—key aspects of EFs. For instance, Lidstone et al. (2010) found that suppressing a child's self-directed speech (SDS) led to poorer performance on the Tower of London task which requires foresight and strategy. TD children increasingly internalize SDS, transitioning from audible to silent and inner self-guidance to support self-regulation and complex planning. DLD can interfere with this transition, potentially stalling the maturation of EFs by hindering the development and internalization of SDS. While there is evidence to support the unidirectional model with EF influencing language and the opposite unidirectional model, research also shows that their interaction is more accurately depicted by a bidirectional relationship (Bruce & Bell, 2022; Salmon et al., 2016). For example, Bohlmann et al. (2015) observed monolingual and dual-language preschoolers over two academic years and found that early language skills could predict later EF, and early EF also predicted later language development. Similarly, Slot and von Suchodoletz (2018) assessed preschoolers' EF components and language abilities over a year. Their analysis showed that not only can early EF predict later language abilities, but early language abilities can also predict later EF performance. Both studies support the notion that language skills and EFs mutually influence each other, underlining the importance of considering both domains in the development of young children.

EFs-narrative relationships

To the best of our knowledge, only one study had reported on the potentially different

involvement of attention and memory skills in narrative macrostructure scores in two different narrative tasks. Specifically, Duinmeijer et al. (2012) found that the macrostructure of narrative recall was uniquely related to memory skills whereas the macrostructure of narrative generation was uniquely related to attention skills in Dutch-speaking school-age children. Intrigued by this finding, we set out to further explore whether different narrative tasks would recruit distinct cognitive skills. We used closely equated narrative task materials and naturalistic parent report of EF skills. Moreover, given the interrelationships among EF, linguistic knowledge, and narrative language (Fiani et al., 2022; Wittke et al., 2013), we controlled for standardized language test scores when evaluating the predictive power of EF components in narrative language skills.

We found that distinct components of EFs were involved in story recall and story generation tasks. For story recall, our results replicated and extended Duinmeijer et al.'s findings (2012) in that WM was the only EF skill that explained unique variance in both macrostructure and microstructure. For story generation, we found that plan/organize, inhibit and WM all predicted macrostructure scores, and inhibit also predicted microstructure scores. Because a mature story model is presented in the story recall task, the most critical cognitive skill required for this task is working memory as children engage in recalling the details in the model. On the other hand, story generation requires a combination of skills. Specifically, planning skills will facilitate the management of the task demands and the integration of word, sentence, and discourse level elements; attention control/inhibition skills will help with sustaining attention on the task and resisting interferences; and working memory will help with monitoring one's own language production and supporting story coherence. The greater EF demands in the story generation task can also explain why young children usually

perform more poorly on this task than on story recall. The informed clinician should be aware that all narrative tasks are not created equal and a seemingly small change in elicitation method may come with a constellation of changes in EF demands.

It is worth noting that we are the first to find a relationship between parent-reported measures of children's EFs and children's narrative production. While it is not our intention to imply a causal relationship between parent-reported EF of their child and their child's narrative performance, we do believe that parents and other familiar adults (e.g., teachers, speech-language pathologists) can integrate verbal scaffolding and feedback to support the development of EF skills. Adults can consciously use guiding questions and comments during daily activities to support the development of children's self-regulation and planning skills. Through interactions with children, adults can not only teach children a broader vocabulary and more complex grammar but can also facilitate the use of self-directed speech which can help children develop EFs. For instance, during a narrative activity, if a child is struggling to develop a story, adults can offer structured feedback that assists the child in organizing their thoughts. Specifically, adults could guide by saying "First, I'll do this... then, I should consider..." By adopting this approach, children are taught to use speech as a tool to navigate their thought processes and actions methodically. This strategy not only enhances the child's ability to construct narratives but also teaches them how to approach tasks systematically, thereby improving their planning skills.

Limitations and future directions

This study has limitations. First, EFs were assessed by parent report, which is prone to informant biases and inaccuracies in retrospective recall. More comprehensive

measurement of EFs should not only collect information about the child's functioning in real-life settings, but also include performance-based assessments. Second, this study was limited by its small sample size. In a study of English-speaking children with and without DLD, Smolak et al (2020) measured visuospatial SSA and WM, and narrative ability in 7-year-olds. Narrative ability was assessed through a standardized test of narrative that aggregates multiple task formats and elicitation methods. They found that SSA was correlated with WM and narrative production in the DLD group, but no significant correlations among these skills were found in the TD group. As in other previous studies (Blom & Boerma, 2016; Duinmeijer et al., 2012), we were not able to run the regressions separately for the DLD and TD groups due to the small sample size. Future studies with sufficient samples of DLD and TD children should aim to examine potential variations in cognition-language relations in impaired and intact systems. In addition, DLD is markedly heterogeneous and future studies with a large sample may examine subsamples of DLD children who are affected by expressive language impairment and mixed receptive-expressive language impairment and query how the language profiles of the children may interact with the cognition-narrative relationship. Third, we assessed EF and narrative language at one time point and found associations between the two sets of skills. This design does not allow us to draw causal conclusions. Future studies may utilize a longitudinal design to better address the direction of the EF-narrative relationship (Blom & Boerma, 2016). Finally, in our comparisons of microstructural measures, we reported the raw numbers for the measures of sentence length and lexical diversity without correcting for differences in sample length, to enable direct comparisons with other narrative studies that reported on these general microstructure measures. A recent study comparing four different lexical diversity measures (Charest et al., 2020) indicated that there are different ways to document

lexical diversity and to correct for narrative length. While moving average type token ratio was the only measure unaffected by sample length, there is not one best measure and different measures may be appropriate for different clinical purposes. In light of these findings, future studies of DLD-TD comparisons may consider utilizing other measures of lexical diversity and sentence length to address the discrepancy in narrative sample length among participant groups.

Conclusion

The current study extends prior work and offers evidence that the act of recalling versus generating stories involve different sets of cognitive skills among Mandarin-speaking preschoolers. Future investigations with a more robust research design should attempt to clarify the nature of these relationships. Furthermore, the findings may motivate future assessment and intervention activities to encompass both sets of skills.

Acknowledgements

We are grateful to all participants who participated in this study.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research was supported by grant 2022YFC2705200 from the National Key Research and Development Program of China and a Hong Kong Research Grants Council General Research Fund Project 15613323 titled "Understanding vulnerabilities in narrative performance in Mandarin-speaking children with developmental language

disorder" to Li Sheng; the JINYE ZHONGZI Project from SCMC-HN (No. JYZZ-HQ-202103); the Sanya Science and Technology Special Fund (No. 2021GXYL41); and the Key Discipline Program of the Sixth Round of the Three-Year Public Health Action Plan (2023-2025) of Shanghai (Grant No. GWVI-11.1-31).

References

- Altman, C., Armon-Lotem, S., Fichman, 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. <https://doi.org/10.1017/S0142716415000466>
- Andreou, G., & LEMONI, G. (2020). Narrative Skills of Monolingual and Bilingual Pre-School and Primary School Children with Developmental Language Disorder (DLD): A Systematic Review. *Open Journal of Modern Linguistics*, *10*(05), 429–458. <https://doi.org/10.4236/ojml.2020.105026>
- Berman, Ruth Aronson & Slobin, Dan Isaac. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, N.J. : Lawrence Erlbaum Associates.
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & and the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. <https://doi.org/10.1111/jcpp.12721>
- Bitetti, D., Hammer, C. S., & López, L. M. (2020). The narrative macrostructure production of Spanish–English bilingual preschoolers: Within- and cross-

- language relations. *Applied Psycholinguistics*, 41(1), 79–106.
<https://doi.org/10.1017/S0142716419000419>
- Blom, E., & Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure? *Research in Developmental Disabilities*, 55, 301–311. <https://doi.org/10.1016/j.ridd.2016.05.001>
- Bohlmann, N. L., Maier, M. F., & Palacios, N. (2015). Bidirectionality in Self-Regulation and Expressive Vocabulary: Comparisons Between Monolingual and Dual Language Learners in Preschool. *Child Development*, 86(4), 1094–1111. <https://doi.org/10.1111/cdev.12375>
- Boudreau, D. (2008). Narrative Abilities: Advances in Research and Implications for Clinical Practice. *Topics in Language Disorders*, 28(2), 99. <https://doi.org/10.1097/01.TLD.0000318932.08807.da>
- Bruce, M., & Bell, M. A. (2022). Vocabulary and Executive Functioning: A Scoping Review of the Unidirectional and Bidirectional Associations across Early Childhood. *Human Development*, 66(3), 167–187. <https://doi.org/10.1159/000524964>
- Charest, M., Skoczylas, M. J., & Schneider, P. (2020). Properties of Lexical Diversity in the Narratives of Children With Typical Language Development and Developmental Language Disorder. *American Journal of Speech-Language Pathology*, 29(4), 1866–1882. https://doi.org/10.1044/2020_AJSLP-19-00176
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: An investigation of their narratives and memory. *International Journal of Language*

- & *Communication Disorders*, 43(2), 201–218.
<https://doi.org/10.1080/13682820701366147>
- Duinmeijer, I., de Jong, J., & Scheper, A. (2012). Narrative abilities, memory and attention in children with a specific language impairment. *International Journal of Language & Communication Disorders*, 47(5), 542–555.
<https://doi.org/10.1111/j.1460-6984.2012.00164.x>
- Durrleman, S., Peristeri, E., & Tsimpli, I. M. (2022). The language-communication divide: Evidence from bilingual children with atypical development. *Evolutionary Linguistic Theory*, 4(1), 5–51.
<https://doi.org/10.1075/elt.00037.dur>
- Ebert, K. D. (2020). Language Sample Analysis With Bilingual Children: Translating Research to Practice. *Topics in Language Disorders*, 40(2), 182.
<https://doi.org/10.1097/TLD.0000000000000209>
- Farrant, B. M., Maybery, M. T., & Fletcher, J. (2012). Language, Cognitive Flexibility, and Explicit False Belief Understanding: Longitudinal Analysis in Typical Development and Specific Language Impairment: Language, Cognitive Flexibility, and False Belief. *Child Development*, 83(1), 223–235.
<https://doi.org/10.1111/j.1467-8624.2011.01681.x>
- Fiani, R., Henry, G., & Prévost, P. (2022). Macrostructure in narratives produced by Lebanese Arabic-French bilingual children: Developmental trends and links with language dominance, exposure to narratives and lexical skills. *Linguistic Approaches to Bilingualism*, 12(4), 446–478.
<https://doi.org/10.1075/lab.20035.fia>
- Gagarina, N. V., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual assessment instrument

- for narratives. *ZAS Papers in Linguistics*, 56, 155.
<https://doi.org/10.21248/zaspil.56.2019.414>
- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, 134(1), 31–60.
<https://doi.org/10.1037/0033-2909.134.1.31>
- Genenbacher, K. (2013). Executive Functions as Rated by Teachers of Preschool Children with Language Impairment and Typical Language Skills. *Undergraduate Honors Theses*. https://thekeep.eiu.edu/honors_theses/70
- Hao, Y., Sheng, L., Zhang, Y., Jiang, F., de Villiers, J., Lee, W., & Liu, X. L. (2018). A Narrative Evaluation of Mandarin-Speaking Children With Language Impairment. *Journal of Speech, Language, and Hearing Research: JSLHR*, 61(2), 345–359. https://doi.org/10.1044/2017_JSLHR-L-16-0367
- Henry, L. A., & Botting, N. (2017). Working memory and developmental language impairments. *Child Language Teaching and Therapy*, 33(1), 19–32.
<https://doi.org/10.1177/0265659016655378>
- Justice, L. M., Bowles, R., Pence, K., & Gosse, C. (2010). A scalable tool for assessing children’s language abilities within a narrative context: The NAP (Narrative Assessment Protocol). *Early Childhood Research Quarterly*, 25(2), 218–234.
<https://doi.org/10.1016/j.ecresq.2009.11.002>
- Kapa, L. L., & Mettler, H. M. (2021). Language and Executive Function in Preschoolers With Developmental Language Disorder: The Role of Self-Directed Speech. *Perspectives of the ASHA Special Interest Groups*, 6(6), 1315–1326.
https://doi.org/10.1044/2021_PERSP-21-00010

- Kapa, L. L., & Plante, E. (2015). Executive Function in SLI: Recent Advances and Future Directions. *Current Developmental Disorders Reports*, 2(3), 245–252.
<https://doi.org/10.1007/s40474-015-0050-x>
- Kapa, L. L., Plante, E., & Doubleday, K. (2017). Applying an Integrative Framework of Executive Function to Preschoolers With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 60(8), 2170–2184.
https://doi.org/10.1044/2017_JSLHR-L-16-0027
- Khan, K. S., Logan, J., Justice, L. M., Bowles, R. P., & Piasta, S. B. (2021). The Contribution of Vocabulary, Grammar, and Phonological Awareness Across a Continuum of Narrative Ability Levels in Young Children. *Journal of Speech, Language, and Hearing Research*, 64(9), 3489–3503.
https://doi.org/10.1044/2021_JSLHR-20-00403
- Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, 37(1), 123–144.
<https://doi.org/10.1017/S0142716415000442>
- Lai xiaochan & wang qiuling. (2017). Narrative Indicators for Effectively Distinguishing between Second-Grade Students with Language Disorders and their Typically Developing Peers. *Bulletin of Special Education*, 42(3), 033–062.
- Lidstone, J. S. M., Meins, E., & Fernyhough, C. (2010). The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. *Journal of Experimental Child Psychology*, 107(4), 438–451.
<https://doi.org/10.1016/j.jecp.2010.06.002>
- Liu, X. L., de Villiers, J., Ning, C., Rolfhus, E., Hutchings, T., Lee, W., Jiang, F., & Zhang, Y. W. (2017). Research to Establish the Validity, Reliability, and

- Clinical Utility of a Comprehensive Language Assessment of Mandarin. *Journal of Speech, Language, and Hearing Research*, 60(3), 592–606.
https://doi.org/10.1044/2016_JSLHR-L-15-0334
- Lu TF, Shuai L, Zhang JS, Wang YF, Qian Y, Zhang HF, Xia WP, Wang ZY, & Tan X. (2017). Validity and reliability of the Behavior Rating Scale of Executive Function-Preschool Version parent form in China. *Chinese Mental Health Journal*, 31(2), 138–143. <https://doi.org/10.3969/j.issn.1000-6729.2017.02.009>
- Luo, J., Yang, W., Chan, A., Cheng, K., Kan, R., & Gagarina, N. (2020). The Multilingual Assessment Instrument for Narratives (MAIN): Adding Mandarin to MAIN. *ZAS Papers in Linguistics*, 64, 159–162.
<https://doi.org/10.21248/zaspil.64.2020.569>
- MacWhinney, B. (2018). *CLAN Manual*. <https://doi.org/10.21415/T5G10R>
- Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, 52(2), 464–488.
<https://doi.org/10.3758/s13428-019-01246-w>
- Merritt, D. D., & Liles, B. Z. (1989). Narrative Analysis. *Journal of Speech and Hearing Disorders*, 54(3), 438–447. <https://doi.org/10.1044/jshd.5403.438>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Ning, C., Liu, X. L., & de Villiers, J. (2014). The diagnostic receptive and expressive assessment of Mandarin. *Dallas, TX: Bethel Hearing and Speaking Training Center*.

- Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2020). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 23(9), 1083–1107.
<https://doi.org/10.1080/13670050.2018.1434124>
- Pauls, L. J., & Archibald, L. M. D. (2016). Executive Functions in Children With Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, 59(5), 1074–1086.
https://doi.org/10.1044/2016_JSLHR-L-15-0174
- Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37(1), 49–67. <https://doi.org/10.1017/S0142716415000417>
- Roello, M., Ferretti, M. L., Colonnello, V., & Levi, G. (2015). When words lead to solutions: Executive function deficits in preschool children with specific language impairment. *Research in Developmental Disabilities*, 37, 216–222.
<https://doi.org/10.1016/j.ridd.2014.11.017>
- Rojas-Barahona, C. A., Förster, C. E., Moreno-Ríos, S., & McClelland, M. M. (2015). Improvement of Working Memory in Preschoolers and Its Impact on Early Literacy Skills: A Study in Deprived Communities of Rural and Urban Areas. *Early Education and Development*, 26(5–6), 871–892.
<https://doi.org/10.1080/10409289.2015.1036346>
- Salmon, K., O’Kearney, R., Reese, E., & Fortune, C.-A. (2016). The Role of Language Skill in Child Psychopathology: Implications for Intervention in the Early Years. *Clinical Child and Family Psychology Review*, 19(4), 352–367.
<https://doi.org/10.1007/s10567-016-0214-1>

- Sheng, L., Shi, H., Wang, D., Hao, Y., & Zheng, L. (2020). Narrative Production in Mandarin-Speaking Children: Effects of Language Ability and Elicitation Method. *Journal of Speech, Language, and Hearing Research*, 63(3), 774–792. https://doi.org/10.1044/2019_JSLHR-19-00087
- Sheng, L., Yu, J., Su, P. L., Wang, D., Lu, T.-H., Shen, L., Hao, Y., & Lam, B. P. W. (2023). Developmental language disorder in Chinese children: A systematic review of research from 1997 to 2022. *Brain and Language*, 241, 105268. <https://doi.org/10.1016/j.bandl.2023.105268>
- Sherman, E. M. S., & Brooks, B. L. (2010). Behavior Rating Inventory of Executive Function – Preschool Version (BRIEF-P): Test Review and Clinical Guidelines for Use. *Child Neuropsychology*, 16(5), 503–519. <https://doi.org/10.1080/09297041003679344>
- Slot, P. L., & von Suchodoletz, A. (2018). Bidirectionality in preschool children’s executive functions and language skills: Is one developing skill the better predictor of the other? *Early Childhood Research Quarterly*, 42, 205–214. <https://doi.org/10.1016/j.ecresq.2017.10.005>
- Smolak, E., McGregor, K. K., Arbisi-Kelm, T., & Eden, N. (2020). Sustained Attention in Developmental Language Disorder and Its Relation to Working Memory and Language. *Journal of Speech, Language, and Hearing Research: JSLHR*, 63(12), 4096–4108. https://doi.org/10.1044/2020_JSLHR-20-00265
- Srivastava, V., Chan, A., & Gagarina, N. (2023). *A comparison of narrative abilities between bilingual children with typical development and bilingual children with developmental language disorder (DLD): A systematic review*. 14th International Symposium on Bilingualism, Sydney, Australia.

- Trainor, K. (2012). Executive functions in preschool children with typical and disordered language. *Masters Theses*. <https://thekeep.eiu.edu/theses/757>
- 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. <https://doi.org/10.1017/S0142716415000478>
- Vallotton, C., & Ayoub, C. (2011). Use your words: The role of language in the development of toddlers' self-regulation. *Early Childhood Research Quarterly*, *26*(2), 169–181. <https://doi.org/10.1016/j.ecresq.2010.09.002>
- Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology*, *141*, 65–82. <https://doi.org/10.1016/j.jecp.2015.06.015>
- Vissers, C., Koolen, S., Hermans, D., Scheper, A., & Knoors, H. (2015). Executive functioning in preschoolers with specific language impairment. *Frontiers in Psychology*, *6*. <https://doi.org/10.3389/fpsyg.2015.01574>
- Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities*, *34*(9), 2586–2597. <https://doi.org/10.1016/j.ridd.2013.05.014>
- Vugs, B., Hendriks, M., Cuperus, J., Knoors, H., & Verhoeven, L. (2017). Developmental Associations Between Working Memory and Language in Children With Specific Language Impairment: A Longitudinal Study. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3284–3294. https://doi.org/10.1044/2017_JSLHR-L-17-0042

- Vugs, B., Hendriks, M., Cuperus, J., & Verhoeven, L. (2014). Working memory performance and executive function behaviors in young children with SLI. *Research in Developmental Disabilities, 35*(1), 62–74.
<https://doi.org/10.1016/j.ridd.2013.10.022>
- Vugs, B., Knoors, H., Cuperus, J., Hendriks, M., & Verhoeven, L. (2016). Interactions between working memory and language in young children with specific language impairment (SLI). *Child Neuropsychology, 22*(8), 955–978.
<https://doi.org/10.1080/09297049.2015.1058348>
- Walker, E. A., Harrison, M., Baumann, R., Moeller, M. P., Sorensen, E., Oleson, J. J., & McCreery, R. W. (2023). Story Generation and Narrative Retells in Children Who Are Hard of Hearing and Hearing Children. *Journal of Speech, Language, and Hearing Research, 66*(9), 3550–3573.
https://doi.org/10.1044/2023_JSLHR-23-00084
- Wechsler, D. (2018). *Wechsler Preschool and Primary Scale of Intelligence—Revised* [Dataset]. <https://doi.org/10.1037/t48859-000>
- Weiland, C., Barata, M. C., & Yoshikawa, H. (2014). The Co-Occurring Development of Executive Function Skills and Receptive Vocabulary in Preschool-Aged Children: A Look at the Direction of the Developmental Pathways. *Infant and Child Development, 23*(1), 4–21. <https://doi.org/10.1002/icd.1829>
- Westerveld, M. F., & Gillon, G. T. (2010). Oral narrative context effects on poor readers' spoken language performance: Story retelling, story generation, and personal narratives. *International Journal of Speech-Language Pathology, 12*(1), 1–11. <https://doi.org/10.3109/17549500903414440>
- Winters, K. L., Jasso, J., Pustejovsky, J. E., & Byrd, C. T. (2022). Investigating Narrative Performance in Children With Developmental Language Disorder: A

Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research: JSLHR*, 65(10), 3908–3929.

https://doi.org/10.1044/2022_JSLHR-22-00017

Wittke, K., Spaulding, T. J., & Schechtman, C. J. (2013). Specific Language Impairment and Executive Functioning: Parent and Teacher Ratings of Behavior. *American Journal of Speech-Language Pathology*, 22(2), 161–172.

[https://doi.org/10.1044/1058-0360\(2012/11-0052\)](https://doi.org/10.1044/1058-0360(2012/11-0052))

Yang, H.-C., & Gray, S. (2017). Executive Function in Preschoolers with Primary Language Impairment. *Journal of Speech, Language, and Hearing Research*, 60(2), 379–392. https://doi.org/10.1044/2016_JSLHR-L-15-0267

Zhu YM, Lu SY, Tang CH, Song J, Gao ES, & Gu XY. (1984). The employment of the Wechsler Pre-school and Primary Scale of Intelligence in urban Shanghai [in Chinese]. *Information on Psychology Science*, 5, 24–31.

Table 1. Background information for children with and without DLD

Table 2. Macrostructure scores

Table 3. Microstructure scores

Table 4. Comparison of BRIEF-P scores between children with and without DLD

Table 5. Association between narrative performance and executive functions

Table 1. Background information for children with and without DLD

	DLD (N=14)	TD (N=34)	χ^2 / t	<i>Effect size</i>
Age (in month)	59.73 (56.63, 67.17)	62.25 (52.83, 65.93)	230.5 ^{&}	-0.03
Gender			0.57 [#]	0.00
male	11 (78.6%)	23 (67.6%)		
female	3 (21.4%)	11 (32.4%)		
Mandarin			2.23 [#]	0.06
<40%	2 (14.3%)	2 (5.9%)		
40-79%	8 (57.1%)	15 (44.1%)		
80-100%	4 (28.6%)	17 (50.0%)		
Mother's			3.36 [#]	0.22
Below bachelor	6 (42.9%)	6 (17.6%)		
Bachelor and	8 (57.1%)	28 (82.4%)		
DREAM-C				
Total Language	84.5 ± 6.21	108.9 ± 8.91	-10.76 [§]	-3.09***
Receptive	87.1 ± 7.88	110.0 ± 8.75	-8.85 [§]	-2.67***
Expressive	71.6 ± 6.58	102.62 ± 12.66	-11.09 [§]	-3.02***
Semantics	91.1 ± 13.0	118.3 ± 11.37	-6.81 [§]	-2.14***
Syntax	77.6 ± 5.33	99.2 ± 7.80	-11.04 [§]	-3.16***
Performance IQ	103.8 ± 9.51	120.1 ± 13.21	-4.79 [§]	-1.38***

***p<0.001.

[&]Non-normally distributed variables were compared using the Mann–Whitney's *U* test and the results were expressed in quartiles [P50 (P25, P75)]. Effect sizes were estimated using rank-biserial correlation (*r*).

[#]Categorical variables were compared using chi-square test and expressed as count (percent). Effect sizes were estimated by Cramer's *V*.

[§]Normally distributed variables were compared using Student t-test and expressed as mean ± SD. Effect sizes were estimated using Hedges' *g*.

Table 2. Macrostructure scores

	DLD (N=14)	TD (N=34)	Effect sizes <i>r</i>
Story generation			
Story structure score	3.00 (2.00, 3.75)	6.00 (4.00, 8.00)	-0.59***
Structural complexity	1.00 (0.00, 1.75)	2.00 (1.00, 2.00)	-0.42*
Internal state terms	2.00 (1.00, 3.00)	4.00 (2.00, 4.00)	-0.38*
Story recall			
Story structure score	6.00 (2.25, 7.00)	9.00 (8.00, 11.00)	-0.66***
Structural complexity	1.50 (0.25, 2.00)	2.00 (1.25, 2.00)	-0.31
Internal state terms	3.00 (0.25, 5.00)	7.00 (5.00, 9.75)	-0.59***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; the results were expressed in quartiles [P50 (P25, P75)]

Table 3. Microstructure scores

	DLD (N=14)	TD (N=34)	Effect sizes <i>r</i>
Story generation			
Mean length of utterance	4.94 (3.97, 5.56)	6.73 (5.44, 7.56)	-0.68***
Total number of words	53.00 (43.00, 71.00)	73.00 (52.00, 85.25)	-0.32
Number of different words	27.50 (20.75, 35.00)	41.50 (30.25, 46.75)	-0.48**
Story recall			
Mean length of utterance	5.45 (4.50, 5.91)	7.13 (6.11, 7.78)	-0.68***
Total number of words	72.50 (44.50, 83.00)	106.50 (76.25, 123.75)	-0.61***
Number of different words	35.50 (24.25, 42.00)	55.0 (40.75, 63.25)	-0.68***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; the results were expressed in quartiles [P50 (P25, P75)]

Table 4. Comparison of BRIEF-P scores between children with and without DLD

	DLD (N=14)	TD (N=34)	<i>W/t</i>	<i>Effect size</i>	<i>p</i>
Inhibit	31.5±7.10	26.5±6.35	2.27 [§]	0.71	0.03*
Shift	16 (13, 18)	12 (11, 15)	340.50 [#]	0.43	0.02*
Emotional control	16.00(13.75, 17.75)	14.5(12, 17)	310.50 [#]	0.30	0.10
Working memory	33.0±8.45	28.1±6.14	1.98 [§]	0.64	0.06
Plan/organize	17.86±4.90	15.6±3.12	1.61 [§]	0.53	0.13
GEC*	114.7±24.72	98.0±18.89	2.28 [§]	0.73	0.03*

* $p < 0.05$

GEC* = Global Executive Composite score

[#] Non-normally distributed variables were compared using the Mann–Whitney’s *U* test and the results were expressed in quartiles [P50 (P25, P75)]. Effect sizes were estimated using rank-biserial correlation (*r*).

[§] Normally distributed variables were compared using Student *t* test and expressed as mean±SD. Effect sizes were estimated using Hedges’ *g*.

Table 1. Association between narrative performance and executive functions

independent variable		Story generation		Story recall	
		β	p	β	p
Story structure score	Inhibit	-0.186	<0.001*	-0.110	0.521
	Working memory	-0.165	0.001*	-0.133	0.021*
	Shift	-0.139	0.184	-0.105	0.387
	Emotional control	-0.197	0.081	-0.148	0.248
	Plan/organize	-0.268	0.004*	-0.174	0.104
Mean length of utterance	Inhibit	-0.052	0.046*	-0.240	0.398
	Working memory	-0.043	0.082	-0.052	0.047*
	Shift	-0.001	0.999	-0.081	0.142
	Emotional control	-0.045	0.414	-0.086	0.139
	Plan/organize	-0.052	0.263	-0.023	0.641
Number of different words	Inhibit	-0.643	0.010*	-0.506	0.095
	Working memory	-0.434	0.072	-0.614	0.031*
	Shift	-0.263	0.603	-0.245	0.685
	Emotional control	-0.837	0.111	-0.633	0.317
	Plan/organize	-0.618	0.167	-0.579	0.279
Total number of words	Inhibit	-1.626	0.006*	-1.205	0.077
	Working memory	-1.103	0.058	-1.407	0.028*
	Shift	-0.194	0.876	-0.243	0.859
	Emotional control	-2.166	0.080	-0.941	0.511
	Plan/organize	-1.806	0.086	-1.166	0.335

All models were adjusted for structural language scores derived from the DREAM-C. * $p < 0.05$.