Taylor & Francis
Taylor & Francis Group

Clinical Linguistics & Phonetics

ISSN: 0269-9206 (Print) 1464-5076 (Online) Journal homepage: https://www.tandfonline.com/loi/iclp20

Does a narrative retelling task improve the assessment of language proficiency in school-aged children born very preterm?

Lottie W. Stipdonk, Jeroen Dudink, Irwin K. Reiss & Marie-Christine J. P. Franken

To cite this article: Lottie W. Stipdonk, Jeroen Dudink, Irwin K. Reiss & Marie-Christine J. P. Franken (2020): Does a narrative retelling task improve the assessment of language proficiency in school-aged children born very preterm?, Clinical Linguistics & Phonetics, DOI: 10.1080/02699206.2020.1720824

To link to this article: https://doi.org/10.1080/02699206.2020.1720824

Published online: 03 Feb 2020.

CLINICAL LINGUISTICS & PHONETICS
https://doi.org/10.1080/02699206.2020.1720824

Does a narrative retelling task improve the assessment of language proficiency in school-aged children born very preterm?

Lottie W. Stipdonk[a], Jeroen Dudink[b,c], Irwin K. Reiss[b], and Marie-Christine J. P. Franken[a]

[a]Department of Otorhinolaryngology, Erasmus Medical University Centre-Sophia Children's Hospital, Rotterdam, Netherlands; [b]Division of Neonatology, Department of Pediatrics, Erasmus Medical University Centre-Sophia Children's Hospital, Rotterdam, Netherlands; [c]Division of Neonatology, Department of Pediatrics, UMCU-Wilhelmina Children's Hospital, Utrecht, Netherlands

ABSTRACT

Almost half of the children born very preterm (VP) experience language difficulties at school-age, specifically with more complex language tasks. Narrative retelling is such a task. Therefore, we explored the value of narrative retelling assessment in school-aged children born VP, compared to item-based language assessment. In 63 children born VP and 30 age-matched full-term (FT) controls Renfrew's Bus Story Test and Clinical Evaluation of Language Fundamentals were assessed. The retelling of the Bus Story was transcribed and language complexity and content measures were analyzed with Computerised Language Analysis software. Narrative outcomes of the VP group were worse than that of the FT group. Group differences were significant for the language complexity measures, but not for the language content measures. However, the mean narrative composite score of the VP group was significantly *better* than their mean item-based language score, while in the FT group the narrative score was *worse* than the item-based score. Significant positive correlations between narrative and item-based language scores were found only in the VP group. In conclusion, in VP children narrative retelling appears to be less sensitive to detecting academic language problems than item-based language assessment. This might be related to the mediating role of attention in item-based tasks, that appears not to affect more spontaneous language tasks such as retelling. Therefore, in school-aged children born VP we recommend using narrative assessment, in addition to item-based assessments, because it is more related to spontaneous language and less sensitive to attention problems.

ARTICLE HISTORY
Received 19 November 2019
Revised 21 January 2020
Accepted 21 January 2020

KEYWORDS
Bus Story Test; CELF; narrative retelling; language disorder; prematurity

Introduction

Nowadays, children born very preterm (VP, <32 weeks) represent 1–2% of all live births in developed countries (Saigal & Doyle, 2008). Since the survival rates of infants born VP have improved over the last decades, the number of children with neurodevelopmental problems during childhood has increased (Goldenberg, Culhane, Iams, & Romero, 2008). Approximately 40% of children born VP without major handicaps have neurodevelopmental disabilities at school age such as learning, behavioural and language problems (Aarnoudse-Moens, Duivenvoorden, Weisglas-Kuperus, Van Goudoever, & Oosterlaan, 2012; Bhutta,

CONTACT Lottie W. Stipdonk l.stipdonk@erasmusmc.nl Department of Otorhinolaryngology, Erasmus University Medical Centre-Sophia Children's Hospital, Wytemaweg 80, Rotterdam, CN 3015, Netherlands

Cleves, Casey, Cradock, & Anand, 2002; Botting, Powls, Cooke, & Marlow, 1997; Elgen et al., 2012; Nguyen et al., 2018; Saigal & Doyle, 2008; van Noort-van der Spek, Franken, & Weisglas-Kuperus, 2012). Regarding language, it has been shown that children born VP experience problems with more complex language tasks and to a lesser extent with simpler language tasks such as a receptive vocabulary test (Stipdonk, Dudink, Utens, Reiss, & Franken, in press; van Noort-van der Spek et al., 2012).

Complex language assessments

Complex language tasks require integration of multiple language components. Usually, an overall complex language score is assessed with a standardised item-based test battery such as the Clinical Evaluation of Language Fundamentals-4 (CELF-4)(Semel, Wiig, & Secord, 2010). This item-based language test battery is based on the sum of subtests, each assessing a specific task such as recalling sentences, following directions or formulating sentences. Each language subtest represents one or a few language competency, such as vocabulary, morphology or syntax. However, there is not one single subtest integrating all language components into an overall language performance outcome. Narratives, on the other hand, are not based on discrete skill testing, but require integration of various cognitive, linguistic and social skills (Renfrew, 1969). Therefore, narrative ability can be assumed to represent spontaneous language performance. It has been described as one of the most "ecologically valid ways" in which to measure communicative competence, both in normal populations and in clinical groups (Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Botting, 2002). A narrative assessment represents the telling or retelling of a fictional or factual story. It provides rich information about linguistic microstructures (e.g. vocabulary, morphology and syntax) as well as macrostructures (such as the organization of events in the plot and coherence in the story)(Jansonius et al., 2014). For clinicians, assessing the child's ability to narrate may be useful since this task may contribute to evaluating how the child's daily communication is affected, and give direction for language therapy (Botting, 2002). Besides, the same authors suggest that relatively subtle language difficulties can be detected on the basis of narratives. Since language difficulties of children born VP vary widely, narrative assessment might be specifically useful to this patient group. In comparison with other discourse-level language, such as conversation and free-play, a narrative retelling task requires language use in a specific context and structure and it elicits more complex syntactic structures (Liles, 1993; Nippold, Hesketh, Duthie, & Mansfield, 2005; Scott & Windsor, 2000; Southwood & Russell, 2004). The Bus Story Test is a narrative retelling assessment tool that contains the most recent norm-references for Dutch school-aged children (Jansonius et al., 2014; Renfrew, 1969). Performance on the Bus Story is supposed to be predictive of future language and literacy performance (Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998).

To our knowledge, so far only two studies used a narrative task in children born VP. Crosbie et al. assessed the Bus Story Test in 15 ten-year-old children born VP and 15 full-term (FT) peers, and showed children born VP to have more utterances with mazes and more disruptions (Crosbie, Holm, Wandschneider, & Hemsley, 2011). However, the children born VP produced a similar story compared to that of their FT peers in terms of content, structure, length of story and complexity. There were neither any group differences in most of the standardised measures on the CELF-4 subtests, Wechsler Intelligence Scale for Children-IV (WISC-IV) and British Picture Vocabulary Scales-II

(BPVS-II)(Crosbie et al., 2011). Smith, DeThorne, Logan, Channell, and Petrill (2014) compared 28 VP born twin pairs to 28 FT born twin pairs at 10 years of age and assessed the Test of Narrative Language (TNL) in combination with four subtests of the CELF-4. The VP twin group performed significantly worse on the item-based standardised tests, but, unexpectedly, not on the narrative assessment. The authors encouraged other researchers to evaluate discourse-level language studies among children born VP and also to look into the influence of attention on standardised test performances.

The recently published European Standards of Care for Newborn Health (EFCNI, 2019) recommended the assessment of language problems not only in the first years of life, but also at school-age. However, there is not yet any evidence-based protocol for the assessment of complex language skills in school-age children born VP. Hence, more research is needed to ascertain how to assess complex language functions in school-aged children born VP. Narrative retelling may refer to spontaneous language performance, required for daily conversations, while item-based language scores might refer to more academic language use. A study that compares in more detail narrative retelling in a sample of VP and FT singleton children with standardised item-based language assessments would contribute to diagnosing language difficulties in children born VP more adequately.

Aims

Therefore, the aim of this study is to explore the added value of assessing narrative retelling ability in school-aged children born VP, compared to item-based language assessment. In other words; does narrative retelling ability provide unique information about the language proficiency in school-aged VP children? We expected children born VP to have worse narrative ability than their FT born peers. Besides, we hypothesized that narrative measures of children born VP as well as FT would be associated with their standardised language test scores. However, we also hypothesised that children born VP would score worse on a narrative assessment than on an item-based language assessment.

Materials and method

Participants

This study was part of a longitudinal cohort study into speech, language and brain development in children born VP. The children had been admitted to the NICU at Erasmus University Medical Centre-Sophia's Children's Hospital in Rotterdam, the Netherlands, between October 2005 and September 2008. Ethical approval has been given by the Medical Ethics Committee of Erasmus University Medical Centre (MEC-2015-591). Parents of participants have given written informed consent for participation and publication. The study inclusion flow-chart is presented in Figure 1. The present study concerns 63 children born VP, at age 10 years (T2). Inclusion criteria were: (1) Born with gestational age of 24–32 weeks; (2) No severe disabilities (i.e. cerebral palsy with GMFCS level >1 or severe vision or hearing disabilities); (3) No congenital abnormalities involving speech organs; (4) Singleton birth; (5) Primary language at home is Dutch.

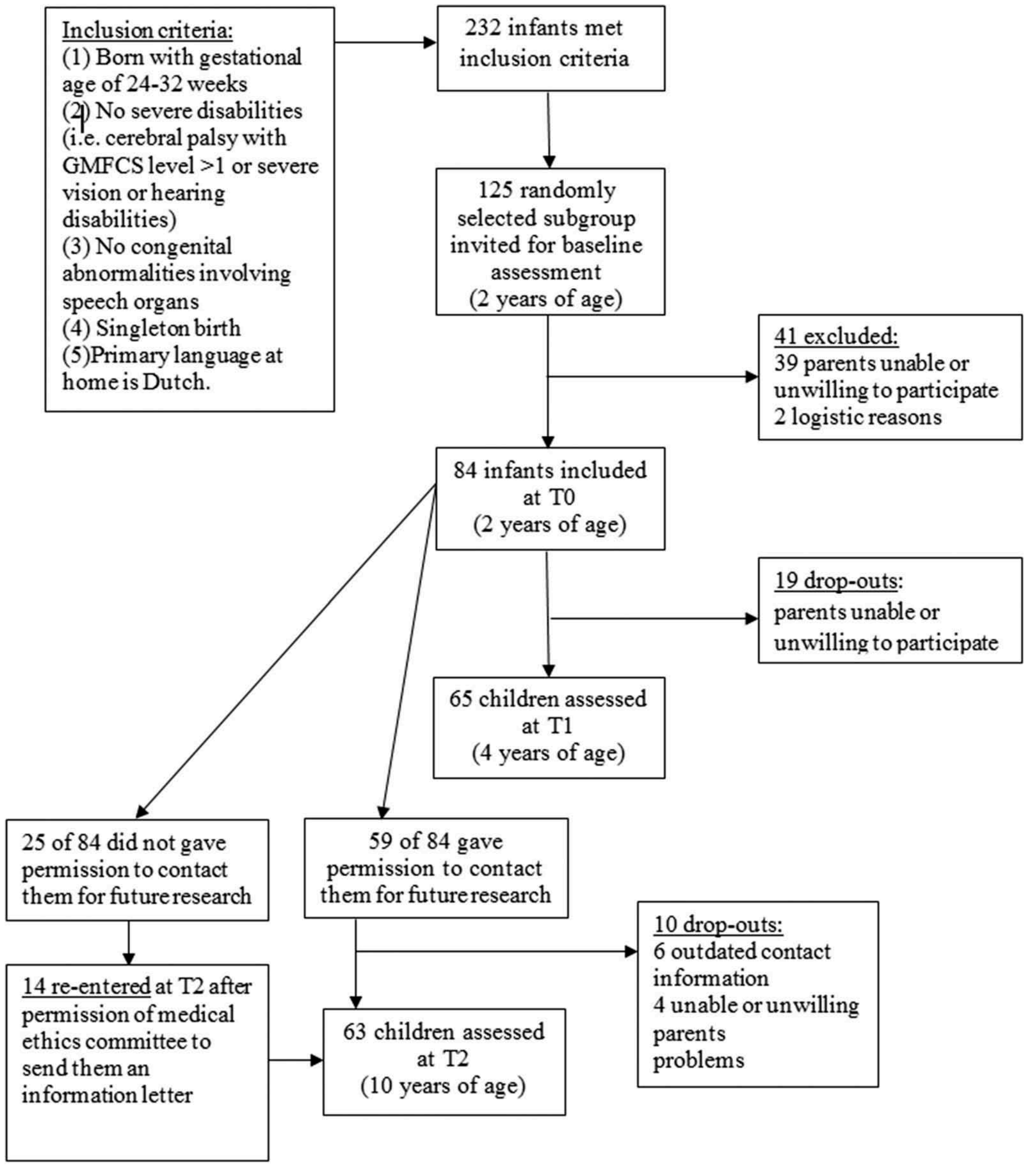

Figure 1. Flow-chart of the inclusion process of the cohort. T0 = baseline time point of the study, at the age of 2. T1 = time point 1, at age 4. T2 = time point 2, at age 10.

As a cross-sectional control group, 30 FT born children, matched on age and sex, were assessed.

Procedure and materials

The Core Language Score of the CELF, the Renfrew Bus Story Test and hearing thresholds were assessed during a one-day visit to Erasmus-MC Sophia's Children's Hospital for both children born VP and FT. In addition, parents of the FT born participants completed a questionnaire requesting: the exact gestational age and birth weight; whether there had been pregnancy or neonatal complications; the educational level of the mother (based on the Dutch educational system); handedness of the child; whether the child had been diagnosed with other disorders (such as ADHD and dyslexia); whether the child had been treated for speech or language difficulties and for how long. This information was already available for the children born VP, since they were being followed longitudinally.

As hearing functioning can affect oral language functions directly, hearing thresholds were measured to detect any hearing losses. A certified clinician according to the ISO standard 8253–1(ISO, 2010) performed pure-tone audiometry (0.5, 1, 2 kHz) and tympanometry in a soundproof booth. A computer-based clinical audiometry system (Decos Technology Group, version 210.2.6 with AudioNigma interface) and TDH-39 headphones were used.

Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4), validated and normed for Dutch children (Semel et al., 2010), is an instrument used to detect language and communication disorders in children of 5–18 years of age (PearsonEducation, 2020). The CELF-4 consists of 11 language subtests. The Core Language Score is the mean score of five of these subtests (i.e. Concepts & Following Directions; Recalling Sentences; Formulating Sentences; Word Classes Receptive and Word Classes Expressive), providing a general language proficiency index. It was administered by a certified speech-language pathologist. Based on a normally distributed scale, the mean standard score for each subtest is 10, and the standard deviation (SD) is 3. The Core Language Score is also normally distributed, however, with a mean of 100 and an SD of 15. Norm references were also converted to z-scores.

The Renfrew Bus Story Test, validated and normed for Dutch children (Jansonius et al., 2014), is an instrument for assessing narrative retelling performance (LearningTools, 2009). Its assessment comprises the retelling of a story about a bus, supported with pictures representing the story, after the story has been read aloud by the examiner in the exact version that is written in the test manual. It was administered by a certified speech-language pathologist. The child's retelling was audio-recorded and transcribed and coded by one of three speech-language pathologists using CHAT (MacWhinney, 2000). Based on these transcriptions, the following outcome measures were determined:

- Information score: The information score indicates the extent to which the child repeated the content of the story correctly.
- Mean Length of Utterances (MLU): The MLU reflects the length of the terminable unit, or T-unit, which refers to a main clause with any subordinate clauses. The MLU provides an index for syntactic development (Nippold et al., 2005; Rice et al., 2010).
- Mean Length of 5 Longest Utterances (ML5LU); The ML5LU provides an index of the complexity of the child's grammatical structures and it represents the maximum language capacity of children better than MLU and is less sensitive to some of the strategies employed to narrate stories, such as using many short sentences (Johnston, 2001; Ketelaars, Jansonius, Cuperus, & Verhoeven, 2016).
- Number of Embedded Utterances (EU): The number of EU indicates clausal density, which is the average number of clauses (main or subordinate) per T-unit and

provides an index of the complexity of the child's grammatical structures (Nippold et al., 2005).

- Number of Ungrammatical Utterances (UU): The number of UU indicates the correctness of utterances and nuances language complexity measures (Liles, Duffy, Merritt, & Purcell, 1995)
- Vocabulary Diversity (VOCD): The VOCD is based on morphological codes of Computerised Language Analysis software CLAN (MacWhinney, 2000). It provides an index of the semantic diversity of the child's language use. In contrast to type-token ratio, VOCD is not impacted by sample size, since it is calculated based on a series of random text samplings. Higher values indicate greater diversity (McCarthy & Jarvis, 2010).

Dutch norm references are available for ML5LU, information score and number of EU of the Bus Story Test for children aged 4 to 10 years. Standard scores were also presented as percentile scores, which we also converted to z-scores. Furthermore, a composite z-score for narrative retelling was calculated, based on the z-scores of these three measures. This overall narrative retelling z-score could be used to compare the score on narrative retelling to the overall item-based language score of the CELF-4.

Reliability

To determine the interrater reliability, one of the three speech-language pathologists also transcribed and analyzed 20% of the samples that had been transcribed by the other two speech-language pathologists. Intraclass correlation coefficient (ICC)(MacWhinney, 2000) and 95% confidence intervals were calculated over the individual scores on the six variables mentioned above. A two-way mixed effects model was used. Between speech-language pathologists 1 and 2, the ICC ranged from .980 to .994 and between speech-language pathologists 1 and 3 the ICC ranged from .984 to .998, which indicates an excellent interrater reliability (Koo & Li, 2016).

Statistics

The statistical analyses were performed using IBM SPSS Statistics version 25. Pearson's chi-square test and independent *t*-test were used to compare the VP children that participated in the present study (n = 63) to the non-participating VP children of the original cohort (n = 169, from total n = 232). Differences in gestational age, birth weight, sex and neighbourhood social economic status were tested. One-way ANOVA and ANCOVA were used to determine the difference between VP and FT children on the narrative measures (information score; MLU; ML5LU; number of EU; number of UU; VOCD; narrative composite score) and the Core Language Score of the CELF, controlled for educational level of the mother, age and sex. A paired samples *t*-test was used to compare mean scores on narrative outcomes to mean Core Language Score for both VP and FT children. The difference between the FT and VP group on the narrative and item-based language difference was measured with ANOVA and ANCOVA. The correlations between the narrative measures and the standardised language scores were measured with

Pearson's correlation coefficients. Differences between the correlation coefficients of the VP and FT group were calculated with Fisher's r to z analysis.

Results

Group characteristics

Gestational age, birth weight, sex and neighborhood social economic status of the study group of children born VP (n = 63) did not significantly differ from the non-participating VP children of the original cohort (n = 169) ($p > .05$; Table 1). Differences between the VP and FT study groups in age at assessment, sex and neighborhood social economic status were also nonsignificant ($p > .05$). However, the difference in the educational level of the mother between VP and FT children approached the level of significance ($p = .051$).

Narrative scores: VP vs FT group

Mean scores and SDs on narrative measures of the VP and FT group and the mean differences between the groups are presented in Table 2. When controlled for educational level of the mother, age and sex, the VP group scored significantly worse on the narrative composite score ($p = .021$), the ML5LU ($p = .012$), the number of EU ($p = .049$) and the item-based language score of the CELF ($p < .000$) than the FT group, based on an ANCOVA. Since the educational level of the mother was missing for five patients, the ANCOVA was based on a patient group of 58. However, ANCOVA results did not differ from ANOVA results based on all 63 patients. The effect sizes were *small to moderate* for

Table 1. Study sample characteristics.

Characteristics	Very preterm (n = 63)	Non-participating Very preterm (n = 169)	Full term (n = 30)
Gestational age in weeks, mean (SD)	29.0 (2.1)	29.3 (1.7)	39.6 (1.3)
Birth weight in grams, mean (SD)	1190 (407)	1217 (338)	3469.1 (450)
Female sex, N (%)	27 (43%)	80 (47%)	11 (37%)
Neighborhood social economic status, mean (SD)	−.04 (.97)	−.02 (.98)	.20 (.82)
Age (years;months) at assessment, mean (SD)	10;6 (0;7)	-	10;3 (0;11)
ADHD diagnosis, N (%)	10 (16%)	-	3 (10%)
Left-handed, N (%)	14 (22%)	-	1 (3%)
Special school services, N(%)	7 (11%)	-	0
Educational level mother, low to high, N (%)	Unknown: 5 (8%)	-	
1: High school	1: 9 (14%)		1: 3 (10%)
2: Secondary vocational education	2: 23 (36%)		2: 5 (17%)
3: Higher vocational education	3: 20 (32%)		3:19 (47%)
4: University level	4: 6 (10%)		4: 10 (26%)
Hearing threshold of one ear above 30 dB – wearing hearing aids	5 (8%) – 3 (5%)	-	0
Hearing threshold of both ears above 30 dB – wearing hearing aids	2 (3%) – 2 (3%)		0
Received speech-language therapy in past	33 (52%)	-	8 (26%)

Table 2. Mean standard scores of narrative measures and Core Language Score of CELF and the composite narrative z-score for VP and FT groups and the effect of group on these measures, based on a one-way ANCOVA, controlled for educational level of the mother, age and sex. Standardised mean-difference effect size (d) was calculated based on means, standard deviations and sample sizes.

	Very preterm *n = 58*		Full term *n = 30*		ANCOVAEffect of group		
ANCOVA	Mean	SD	Mean	SD	F	*p*-value	Effect size (d)
Narrative Composite score	−.37	.86	.04	.64	6.0	.016	.52
Information score	21.5	5.0	22.6	5.0	1.8	.179	.22
ML5LU (in words)	11.7	2.3	12.9	1.8	6.6	.012	.56
Number of Embedded Utterances	3.7	2.3	4.6	2.3	4.0	.049	.39
Number of Ungrammatical Utterances	2.5	1.7	2.0	1.5	1.7	.202	.30
VOCD	37.3	8.2	39.2	6.6	1.0	.311	.25
Core score CELF	89.8	15.7	105.1	11.5	18.1	.000	1.06

the narrative measures and *large* for the Core Language Score of the CELF. No group differences were found on the information score ($p = .179$), number of UU ($p = .220$) nor VOCD ($p = .311$).

Narrative versus item-based language measures

In the VP group, the mean composite z-score of the narrative assessment was significantly *higher* than the mean item-based language z-score of the CELF ($p = .016$). Conversely, in the FT group, the mean narrative composite z-score was *lower* than their item-based language z-score, although this difference did not reach the level of significance ($p = .115$; Figure 2). Consequently, the VP and FT group differed significantly on the difference score between the composite narrative score and item-based language score ($p = .007$, effect size = .62). However, after controlling for the educational level of the mother, age

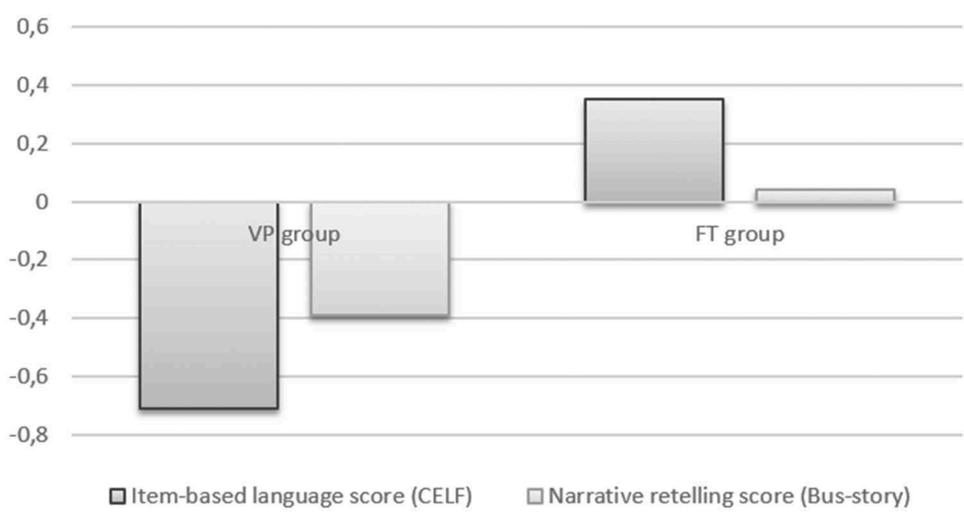

Figure 2. Mean z-scores for Core Language Score of CELF and composite score of narrative retelling of bus-story for VP and FT group.

and sex, the *p*-value of this effect of group (i.e. VP or FT) was $p = .051$ (Table 3). In addition, the number of children born VP with below-average scores (i.e. < −1 SD below the mean of the norm reference group) on the item-based language score of the CELF, in combination with average scores on the narrative composite score was significantly higher (n = 15, 24%) than the number of FT children with this combination of scores (n = 1, 3%; Table 4).

Association measures

All correlations between subtests and item-based language scores of the CELF narrative measures are presented in Appendix A. Scatterplots of the significant associations between narrative and item-based language measures are presented in Figure 3. In the VP group, a significant positive correlation between the item-based language score of the CELF and the narrative information score was found (r = .435, $p < .001$), which was not found in the FT group (r = .135, $p = .477$). Based on Fisher's r to z transformation, these correlation coefficients of the VP and FT group were not statistically significant (z = 1.425, $p = .154$). Between the item-based language score and ML5LU a significant positive correlation was found in the VP group (r = .374, $p = .003$), while in the FT group a negative correlation was found, which,

Table 3. Effect of group in differences on the narrative-language difference of each group, based on ANOVA and ANCOVA, controlled for educational level of the mother, age and sex. Standardised mean-difference effect size (d) is calculated based on means, standard deviations and sample sizes.

	Difference score VP *n* = 63		Difference score FT *n* = 30					
ANOVA	Mean	SD	Mean	SD	Mean diff	95% confidence interval	p-value ANOVA	Effect size (d)
	.31	1.0	−.31	1.0	.62	.20 to 1.09	.007	.62
	Difference score VP *n* = 58		Difference score FT *n* = 30		ANCOVA (N = 58) Effect of group			
ANCOVA	Mean	SD	Mean	SD	F	*p*-value		
	.31	1.1	−.31	1.0	3.9	.051		

Table 4. Number of children scoring above and below −1 SD (i.e. "average" and "below average") on composite narrative score and Core Language Score CELF in VP and FT group.

	Below Average Core Language Score CELF	Average Core Language Score CELF	Total
Below Average Composite narrative score	VP: 13 (20%) FT: 0	VP: 6 (10%) FT: 2 (7%)	VP: 18 (29%) FT: 2 (7%)
Average Composite narrative score	VP: 15 (24%) FT: 1 (3%)	VP: 29 (46%) FT: 27 (90%)	VP: 44 (71%) FT: 27 (93%)
Total	VP: 27 (44%) FT: 1 (3%)	VP: 35 (56%) FT: 29 (97%)	VP: 63 (100%) FT: 30 (100%)

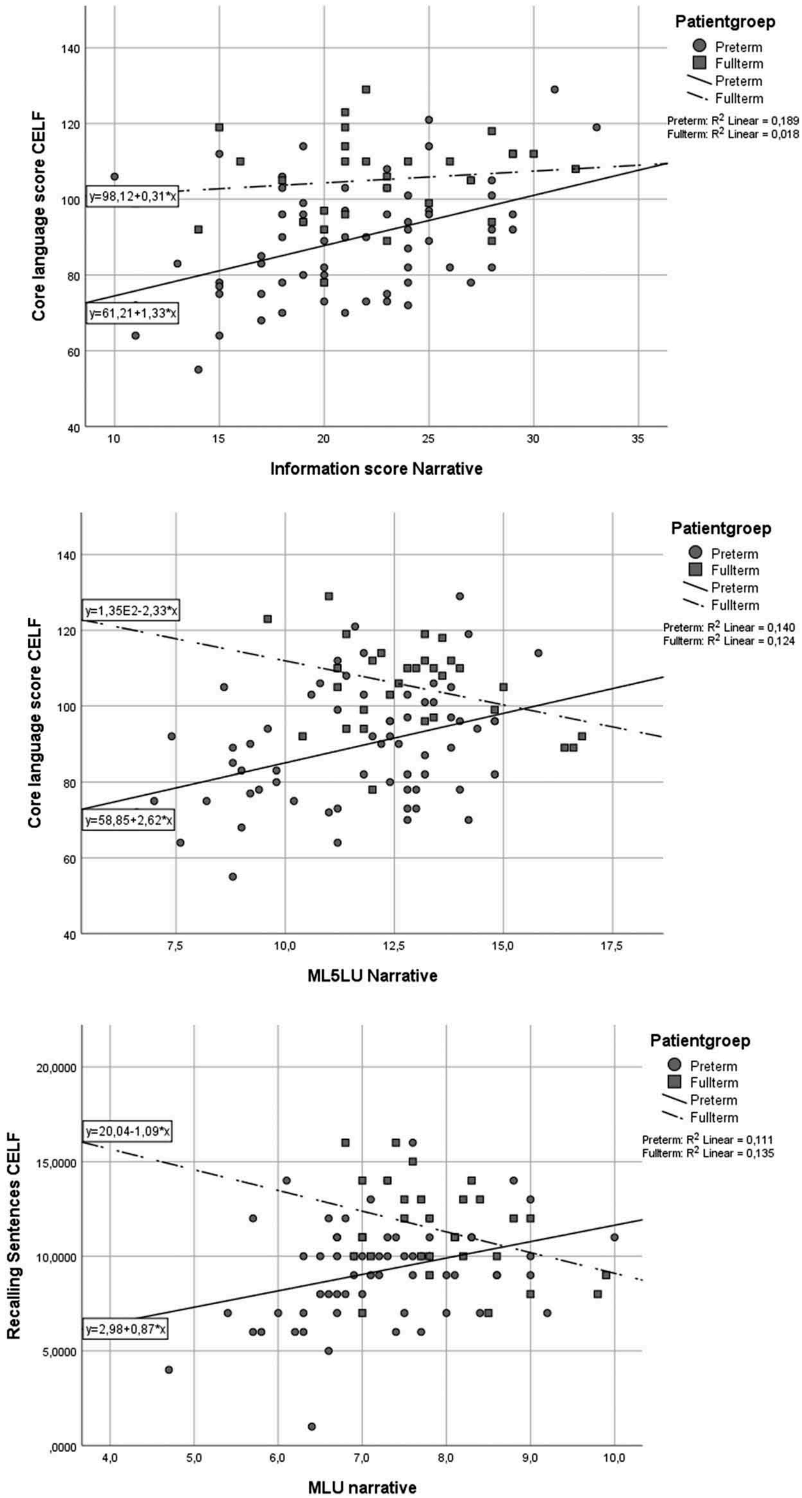

Figure 3. Scatterplots and linear fit lines of VP and FT group.

however, did not reach the level of significance (r = −.353, p = .056). Comparing these correlations, a significant difference was found (z = 3.288, p = .001). The difference between groups on the correlation between the subtest score on Recalling Sentences of the CELF and the MLU score of the narrative retelling task was also significant (z = 3.131, p = .002); the VP group showed a significant positive association (r = .327, p = .009), while the FT group showed a significant negative association (r = −.368, p = .046).

Discussion

Although children born VP without major handicaps score worse on narrative retelling than FT peers, their narrative ability was significantly better than their item-based language skills. FT children, conversely, had worse narrative ability compared to their standardised language skills. More than a quarter of the VP group showed sufficient narrative ability, but below-average scores on an item-based language test. Therefore, our hypothesis (assuming that children born VP experience more problems with narrative retelling than with item-based language assessments since it is technically a more complex task) has to be rejected. Our findings suggest that children born VP have fewer problems with the spontaneous use of language in a narrative retelling task than with the abstract assessment of isolated language skills. Narrative retelling assessment therefore appears to be less sensitive than the assessment of standardised isolated complex language skills in detecting the more academic language difficulties in children born VP. However, the specific associations between narrative measures and item-based language measures that were found only in children born VP showed the added value of narrative retelling assessment in defining and specifying language difficulties in this patient group. Narratives provide detailed information about the type of language difficulties and coping strategies of children born VP. In our group of children born FT, on the other hand, narrative ability was relatively weak. This suggests that language interventions for FT children with language difficulties might need to be more focused on narrative ability than on isolated language skills.

Interpretation and meaning of results

An explanation for the better narrative performance of children born VP might be found in the nature of narratives. Since a narrative is a relatively natural language task, representing the spontaneous use of language more adequately than abstract subtests of an item-based language test, children born VP might experience less difficulties with it.

This difference between tasks might be impacted by the required level of sustained attention in each task (Mahurin-Smith, DeThorne, & Petrill, 2017). It is well-known that children born VP have more attention problems than FT born peers (Aarnoudse-Moens, Weisglas-Kuperus, van Goudoever, & Oosterlaan, 2009; Botting et al., 1997; Elgen, Sommerfelt, & Markestad, 2002). The duration of an item-based language assessment is much longer than that of a narrative retelling assessment, resulting in different levels of sustained attention. Besides, the Bus Story Test is supported with pictures, which might make it easier to concentrate on the task, compared to the numerous items and turn-taking shifts that are required in an item-based language assessment. Thus, a narrative retelling assessment requires less sustained attention than an item-based language assessment. Note that spontaneous narrative telling requires even less sustained attention than a retelling assessment. Following this reasoning, item-based language assessment might overestimate language problems in children with attention problems. In these children, item-based language scores may predominantly reflect academic language functions rather than spontaneous language proficiency. Future research will be needed to explore this idea. Nevertheless, a narrative task may be a valuable addition to item-based language tests, as a task that is more strongly related to spontaneous speech and less sensitive to attention. Consequently, narrative assessment may improve the diagnosis of language difficulties in children born VP.

Furthermore, the relatively good performance on narrative retelling might also be associated with the relatively high vocabulary scores of this group (Stipdonk et al., in press; van Noort-van der Spek et al., 2012). Stipdonk et al. showed in the same study group that mean vocabulary scores were significantly higher than mean scores on the CELF-4-NL (Stipdonk et al., in press). Since narrative retelling ability, in general, is related to vocabulary knowledge, this might be an important association in children born VP.

A more fundamental explanation for our findings might be the atypical language tracts/pathways in the brain of children born VP. Recently, Bruckert et al. (2019) found associations between reading ability and white matter pathways in children born FT, but not in children born VP, which suggested that children born VP might have a larger, but less specific network of white matter pathways involved in reading (Bruckert et al., 2019). If the atypical brain development of children born VP is indeed characterised by a more dispersed network without specifically good language subtracts, this might also explain their weak performance on isolated, specific language tasks and their relatively good performance on, more natural and free, language tasks.

Clinical implications

For clinical purposes, we recommend using narrative retelling assessment, in combination with an attention task or questionnaire and item-based language tasks, in school-aged children born VP. Since retelling and item-based language functioning differed significantly in our VP group, and both skills are needed for adequate language performance, it is relevant to assess both in clinical practice. In combination with attention skill assessment, narrative retelling will be relevant and complementary to item-based standardised language assessment in this patient group.

Narratives in children born VP and FT: agreement with the literature

Although the narrative performance of the VP group was better than their isolated language skills, most of the narrative outcomes were still significantly worse than those of age-matched FT born peers. Specifically, the VP group scored worse than the FT group on measures of the grammatical complexity of their story (i.e. ML5LU and number of EU), but not on content measures (i.e. information score and VOCD). This suggests that children born VP experience more difficulty with using complex grammatical structures in a story than with applying more complex semantics. This result is not entirely in agreement with the results of Crosbie et al. (2011) and Smith et al. (2014) who also assessed narratives in children born VP (Crosbie et al., 2011; Smith et al., 2014). Neither of these studies found any differences between children born VP and FT in the content or complexity of their stories. However, in both studies, the VP group did score significantly worse than the FT group on the subtests of the CELF, which is in accordance with the present study. Since Crosbie et al. studied a relatively small sample size (15 VP and 15 FT children) and Smith et al. studied only twins, we think the present study adds to the literature. It leads to growing evidence that children born VP with attention problems, may have fewer problems with retelling than with item-based language testing.

Associations between narrative and item-based language measures

The positive correlations between narrative measures (Bus Story's ML5LU and Information score) and the Core Language Score of the CELF in the VP group may reflect that children born VP with better language scores use relatively lengthy sentences. This language style might match a low score on one of the five dimensions of language fluency, defined by Fillmore (1979) as "succinctness, the ability to speak in logically organised and semantically dense sentences such that ideas are expressed in a compact and careful way"(in Logan, 2015). In children born FT, on the other hand, the narrative and item-based language scores were independent of the length or complexity of their utterances. Even negative associations were found, suggesting that better standardised language functions were associated with the use of shorter sentences in retelling for children born FT. This is the opposite relationship of that in children born VP. The causality of the relation in the VP group remains unclear, however. In children born VP, there might be a common neurological cause for their language difficulties. As described in the previous paragraph, children born VP might have a more dispersed language network in the brain, which might cause problems with isolated language skills and narrative ability. Another explanation of the significant association might be that language functions of children born VP are influenced by their talking experience, and that externalising talkers develop better language skills than internalising talkers. A third possibility is that children born VP with better language skills feel an urge to perform, and therefore use longer and more complex sentences than children born VP with weaker language skills. More research and longitudinal studies with detailed linguistic analyses are needed to improve our understanding of these associations. Furthermore, it would be interesting to study in more detail the macrostructures of narratives in children born VP.

Strengths and limitations

The strength of this study is that children born VP and age-matched controls were linguistically analyzed in detail; item-based language assessments were performed and transcripts of a narrative retelling task were analyzed. Although analysis of narratives is time-consuming, we studied a relatively big sample with sufficient statistical power. Since the existing literature about language development in children born VP is mainly based on item-based language assessments, this study adds to what was already known on this topic. Our sample seems to be representative for school-aged children born VP without major handicaps, since our sample did not significantly differ from the non-participating group of VP children of our cohort on gestational age, birth weight, sex and social economic status. Yet, our results cannot be generalised for children born late preterm or for other age groups. A limitation of the present study is that the VP and FT group differed significantly on the educational level of the mother. We therefore controlled for this factor in all relevant analyses, together with age and sex. Another possible limitation of this study is that the children born FT might have been less motivated for the assessment than the children born VP. The children born VP were the subject of the study and originally hospital patients, who were assessed for clinical reasons in the past. Therefore, they might have felt more pressure to perform well than the FT children who had no relation to the hospital at all. However, we do not know whether this difference in the clinical record has impacted their test motivation and results. If they have been underperforming, this would mean that their actual narrative ability would be better, which would make the differences with the VP group even bigger. In addition, we think it is improbable that the motivation for the narrative assessment was different from that of the item-based language assessment. Therefore, it appears unlikely that the results on association measures were impacted by their motivation. Another limitation of the study is that the norm references of the Renfrew Bus Story Test are based on children up to 10;0 years of age and that these norm references are relatively old (i.e. standardization studies took place between 2006 and 2013)(Jansonius et al., 2014). The mean age of the children of our study group was 10;6, with a minimum age of 9;0 and a maximum age of 12;0. All children are compared to the norm group of children aged 9;0–10;0. Since we were specifically interested in the differences between our VP and FT group, we do not think this impacted our study results significantly. Besides, it was expected, based on the stabilization of scores of the norm references at 9–10 years of age, that narrative functions no longer develop quickly the age of 10 (Jansonius et al., 2014). However, if it would have impacted our results, the number of children scoring below 1 SD would have been bigger, suggesting that the actual narrative retelling performance of children born VP and FT is worse.

Conclusions

The narrative retelling ability of children born VP is relatively good compared to their standardised language scores, suggesting that children born VP experience fewer problems with language tasks that are more strongly related to the spontaneous use of language, than with item-based assessments of isolated complex language skills. This difference between language tasks might be mediated by attention skills, suggesting that item-based language assessments sometimes overestimate spontaneous language functions in school-aged children born VP. Besides, children born VP with higher language scores tended to produce longer utterances, while the language scores of FT children were independent of their retelling skills. Thus, narrative retelling assessment appears not to be more sensitive to language difficulties than item-based, standardised language tests, but it provides detailed information about the type of language difficulties and coping strategies of children born VP and it may be a more attention-independent language assessment. Adding a narrative retelling assessment to item-based standardised language assessments in school-aged children born VP is, therefore, recommended.

Acknowledgments

We thank all children and their parents for their participation in the study. We thank Darlene Keydeniers and Margriet van der Spek for their contribution to data collection and transcription of the narrative assessment of a subgroup of the patients. We are also grateful for the contribution of Sheean Spoel of the Digital Humanities Lab of Utrecht University for his work on the Dutch parser for the CHAT files and the morphological annotations. Lastly, we thank Woody Starkweather for linguistic support.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Dr. C.J. Vaillantfonds; Stichting Mitialto; Stichting Coolsingel under Grant number [496]; and Stichting Sophia Wetenschappelijk Onderzoek under Grant number [S19-24].

References

Aarnoudse-Moens, C. S., Duivenvoorden, H. J., Weisglas-Kuperus, N., Van Goudoever, J. B., & Oosterlaan, J. (2012). The profile of executive function in very preterm children at 4 to 12 years. *Developmental Medicine & Child Neurology*, *54*(3), 247–253. doi:10.1111/j.1469-8749.2011.04150.x

Aarnoudse-Moens, C. S., Weisglas-Kuperus, N., van Goudoever, J. B., & Oosterlaan, J. (2009). Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics*, *124*(2), 717–728. peds.2008-2816 [pii]. doi:10.1542/peds.2008-2816

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm. *JAMA: the Journal of the American Medical Association*, *288*(6), 728–737. doi:10.1001/jama.288.6.728

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *International Journal of Language & Communication Disorders*, *51*(6), 626–638. doi:10.1111/1460-6984.12234

Botting, N. (2002). Narrative as a tool for the assessment of linguitic and pragmatic impairments. *Child Language Teaching and Therapy*, *18*(1), 1–21. doi:10.1191/0265659002ct224oa

Botting, N., Powls, A., Cooke, R. W., & Marlow, N. (1997). Attention deficit hyperactivity disorders and other psychiatric outcomes in very low birthweight children at 12 years. *Journal of Child Psychology and Psychiatry*, *38*(8), 931–941. doi:10.1111/jcpp.1997.38.issue-8

Bruckert, L., Borchers, L. R., Dodson, C. K., Marchman, V. A., Travis, K. E., Ben-Shachar, M., & Feldman, H. M. (2019). White matter plasticity in reading-related pathways differs in children born preterm and at term: A longitudinal analysis. *Frontiers in Human Neuroscience*, *13*, 139. doi:10.3389/fnhum.2019.00139

Crosbie, S., Holm, A., Wandschneider, S., & Hemsley, G. (2011). Narrative skills of children born preterm. *International Journal of Language & Communication Disorders*, *46*(1), 83–94. doi:10.3109/13682821003624998

(EFCNI), E. F. f. t. C. o. N. I. (2019). *European standards of care for newborn health*. Retrieved from https://newborn-health-standards.org/standards/followup-continuing-care/overview/

Elgen, I., Sommerfelt, K., & Markestad, T. (2002). Population based, controlled study of behavioural problems and psychiatric disorders in low birthweight children at 11 years of age. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *87*(2), F128–132. doi:10.1136/fn.87.2.F128

Elgen, S. K., Leversen, K. T., Grundt, J. H., Hurum, J., Sundby, A. B., Elgen, I. B., & Markestad, T. (2012). Mental health at 5 years among children born extremely preterm: A national population-based study. *European Child & Adolescent Psychiatry*, *21*(10), 583–589. doi:10.1007/s00787-012-0298-1

Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *The Lancet*, *371*(9606), 75–84. S0140-6736(08)60074-4 [pii]. doi:10.1016/S0140-6736(08)60074-4

ISO. (2010). Retrieved from https://www.iso.org/standards.html

Jansonius, K., Ketelaars, M., Bogers., M., Van den Heuvel, E., Roevers, H., Manders, E., & Zink, I. (2014). *Renfrew taalschalen Nederlandse Aanpassing*. Antwerp, Belgium: Garant.

Johnston, J. R. (2001). An alternate MLU calculation: Magnitude and variability of effects. *Journal of Speech, Language, and Hearing Research*, *44*(1), 156–164. doi:10.1044/1092-4388(2001/014)

Ketelaars, M. P., Jansonius, K., Cuperus, J., & Verhoeven, L. (2016). Narrative competence in children with pragmatic language impairment: A longitudinal study. *International Journal of Language & Communication Disorders*, *51*(2), 162–173. doi:10.1111/1460-6984.12195

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. doi:10.1016/j.jcm.2016.02.012S1556-3707(16)00015-8[pii]

LearningTools. (2009). Retrieved from https://www.learningtoolsllc.com/store/p1/Renfrew_Bus_Story_-_North_American_%28RBS-NA%29_edition_.html

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, Language, and Hearing Research*, *36*(5), 868–882. doi:10.1044/jshr.3605.868

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, Language, and Hearing Research*, *38*(2), 415–425. doi:10.1044/jshr.3802.415

Logan, K. J. (2015). *Fluency disorders*. San Diego, CA: Plural Publishing.

MacWhinney, B. (2000). *The Childes project: Tools for analysing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associetes.

Mahurin-Smith, J., DeThorne, L. S., & Petrill, S. A. (2017). Longitudinal associations across prematurity, attention, and language in school-age children. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3601–3608. 2665790 [pii]. doi:10.1044/2017_JSLHR-L-17-0015

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. 42/2/381 [pii]. doi:10.3758/BRM.42.2.381

Nguyen, T. N., Spencer-Smith, M., Zannino, D., Burnett, A., Scratch, S. E., Pascoe, L., & Anderson, P. J. (2018). Developmental trajectory of language from 2 to 13 years in children born very preterm. *Pediatrics*, *141*(5). peds.2017-2831 [pii]. doi:10.1542/peds.2017-2831

Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research*, *48*(5), 1048–1064. doi:10.1044/1092-4388(2005/073)

PearsonEducation. (2020). Retrieved from https://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildLanguage/ClinicalEvaluationofLanguageFundamentals-FourthEditionUK(CELF-4UK)/ClinicalEvaluationofLanguageFundamentals-FourthEditionUK(CELF-4UK).aspx

Renfrew, C. E. (1969). *The Bus Story: A test of continuous speech*. Headington, UK: North Place.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research*, *53*(2), 333–349. 53/2/333 [pii]. doi:10.1044/1092-4388(2009/08-0183)

Saigal, S., & Doyle, L. W. (2008). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*, *371*(9608), 261–269. S0140-6736(08)60136-1 [pii]. doi:10.1016/S0140-6736(08)60136-1

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, *43*(2), 324–339. doi:10.1044/jslhr.4302.324

Semel, E., Wiig, E. H., & Secord, W. A. (2010). *Clinical evaluation of language fundamentals 4 - NL*. Amsterdam, The Netherlands: Pearson.

Smith, J. M., DeThorne, L. S., Logan, J. A., Channell, R. W., & Petrill, S. A. (2014). Impact of prematurity on language skills at school age. *Journal of Speech, Language, and Hearing Research*, *57*(3), 901–916. 1092-4388_2013_12-0347 [pii]. doi:10.1044/1092-4388(2013/12-0347)

Southwood, F., & Russell, A. F. (2004). Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech, Language, and Hearing Research*, *47*(2), 366–376. doi:10.1044/1092-4388(2004/030)

Stipdonk, L. W., Dudink, J., Utens, E. M. W. J., Reiss, I. K., & Franken, M. C. J. P. (in press). Language functions deserve more attention in follow-up of children born very preterm. *European Journal of Paeediatric Neurology*.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired Preschoolers, A follow-up into adolescence. *Journal of Speech, Language and Hearing Research*, *41*(2), 407–418. doi:10.1044/jslhr.4102.407

van Noort-van der Spek, I. L., Franken, M., & Weisglas-Kuperus, N. (2012). Language functions in preterm-born children: A systematic review and meta-analysis. [Review]. *Pediatrics*, *129*(4), 745–754. doi:10.1542/peds.2011-1728

Appendix A

Table A1. Correlation matrix between Standardised language scores of Core Language Score and subtest scores of the CELF and narrative measures.

	Core Language Score		Concepts & Following Directions		Recalling Sentences		Formulating Sentences		Word Classes Receptive		Word Classes Expressive		Word Classes Total	
	VP	FT	VP	FT	VP	FT	VP	FT	VP	FT	VP	FT	VP	FT
Composite narrative score	.378**	.121	.324*	.163	.462**	−.079	.244	.139	.198	.070	.243	.233	.225	.147
Information score	.435**	.135	.372**	.017	.564**	.244	.260*	.091	.230	−.022	.253*	.058	.258*	.016
N Embedded Utterances	.257*	.044	.213	.104	.326**	−.174	.172	.067	.134	.117	.178	.234	.162	.145
ML5LU	.374**	−.353	.316*	−.111	.399**	−.308	.267*	−.343	.227	−.261	.284*	−.083	.249*	−.207
MLU	.314*	−.354	.253*	−.298	.301*	−.368*	.275*	−.288	.204	−.121	.216	.083	.211	−.042
N Ungrammatical Utterances	−.165	−.284	−.169	−.507*	−.072	−.310	−.243	−.114	.051	.108	−.111	.134	−.035	.116
VOCD	.530**	−.205	.332*	.103	.578**	−.217	.462**	−.297	.396*	−.137	.377*	−.070	.397	−.205

** *p*-level = <.001, *p-level = <.05.