

The concurrent and longitudinal relationship between narrative skills and other language skills in children

First Language

1–18

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0142723721995688

journals.sagepub.com/home/fla



Jannicke Karlsen

Department of Special Needs Education, University of Oslo, Norway; Statped, Norway

Hanne Næss Hjetland

Department of Special Needs Education

Bente Eriksen Hagtvet

Department of Special Needs Education

Johan Braeken

Center for Educational Measurement (CEMO)

Monica Melby-Lervåg 

Department of Special Needs Education, University of Oslo, Norway

Abstract

We examined the concurrent relationship between narrative skills (the Renfrew Bus Story Test) and core language measures (vocabulary, grammar and verbal memory) at age 4 and the longitudinal relationship between core language and listening comprehension skills at age 7 in a sample of 215 children using latent variables and structural equation modelling. Our main purpose was to investigate to what extent narrative retell constitutes a unique influence on later language and listening comprehension skills. The results support a two-factor model of narrative retelling and core language representing different but related constructs at age 4. Narrative retell explained unique variance in

Corresponding author:

Monica Melby-Lervåg, Department of Special Needs Education, University of Oslo, 0316 Oslo, Norway.

Email: monica.melby-lervag@isp.uio.no

later language skills but did not explain additional variance beyond the 58% explained by the age 4 language construct. Similarly, narrative retell predicted unique variance in later listening comprehension, but not beyond what was explained by core language skills at age 4. The strength of the relationship between narrative retelling at age 4 and the age 7 measures was not related to the level of narrative skills. The results indicate that age 4 traditional core language measures capture more of the skills that are important for later language and listening comprehension than narrative skills at the same age.

Keywords

Narrative skills, language development, language problems, longitudinal, oral language

For a child, to be able to tell stories and describe events so that peers or adults can understand them has great value for interaction and communication. This is often referred to as having narrative skills and involves how to verbally describe the content of – and the connection between – events (Berman, 1988, p. 470). Most children master narrative skills rather effortlessly; therefore, it is easy to forget that these skills are quite complex and arguably involve a range of language and cognitive skills such as grammar and memory (Bishop & Edmundson, 1987; Norbury et al., 2014; Paul & Smith, 1993). Furthermore, storytelling appears to develop interdependently, thereby influencing other language and literacy skills (Bishop & Edmundson, 1987; Stothard et al., 1998), including higher order language skills such as listening and reading comprehension (e.g. Cain, 2003; Kendeou et al., 2009; Kim et al., 2015; Mäkinen et al., 2018; Suggate et al., 2018). The claim has further been made that narrative skills are a more sensitive marker of persistent language problems than are other language measures, such as vocabulary or grammar (Bishop & Edmundson, 1987; Pankratz et al., 2007; Vandewalle et al., 2012). Also, intervention studies have demonstrated that narrative skills, in contrast to vocabulary for instance, seem highly malleable in interventions and comprise a sensitive measure to capture intervention effects (Rogde et al., 2019). In short, narrative skills appear to play a crucial, but thus far little understood, role in language development. Thus, longitudinal studies of the development of narrative production in relation to other areas, such as core language skills, have been recommended as a future area of research (Norbury et al., 2014).

In this study, we examine to what extent narrative skills measured using a frequently used narrative retell task (the Bus Story Test) relate to and predict other language skills, both concurrently and longitudinally, and whether the ability to retell an orally presented story is a stronger predictor of later language skills in those with weaker language skills than in those with better language skills. Clarifying these issues will not only offer valuable information about developmental relations between different language skills but also shed light on assessment and the structure of language skills in young children, including whether narrative skills overlap with core language skills (i.e. grammar and vocabulary) so that they effectively comprise the same construct. Finally, this examination will reveal information that can be tested in intervention studies, that is, whether narrative skills are best targeted directly on their own or together with other skills.

Narrative language skills and the dimensionality of language

Oral language skills have often been considered multidimensional and to consist of different components, such as vocabulary, syntax and semantics. However, an increasing number of studies indicate that rather than consisting of different but related skills, oral language is better understood as unidimensional. This is supported in both preschool-aged children (Bornstein & Putnick, 2012; Colledge et al., 2002) and children up to fifth grade (Kendeou et al., 2009; Protopapas et al., 2012; Tomblin & Zhang, 2006; Storch & Whitehurst, 2002). Particularly, it appears that discourse-level language skills such as narrative proficiency or listening comprehension are differentiated from core language skills (e.g. vocabulary and grammar), at least in older children (e.g. Language and Reading Research Consortium [LARRC], 2015; Tomblin & Zhang, 2006).

Discourse-level skills such as narrative production involve a range of skills, including the understanding of stories and mastery of pragmatics, syntax and semantics (Suggate et al., 2018). Previous studies have found relationships between various aspects of narrative proficiency and children's comprehension (Cain, 2003; Kim et al., 2005; Mäkinen et al., 2018), indicating that measures of narrative retell and comprehension capture common discourse-level skills (Kim et al., 2015).

The complexity of narrative proficiency is often understood at two levels: the ability to organize a story at the macro level and linguistic skills at the micro level. These two levels of narrative organization are often referred to as *story coherence* and *linguistic cohesion*. Story coherence, or macro-structure, refers to how the events are related to each other and to the overall theme of the story. Linguistic coherence, or micro-structure, describes the semantic relations between sentences, often with a focus on the words used and the linguistic complexity of the stories. Both levels are related to comprehension (including reading comprehension, e.g. Cain, 2003; Mäkinen et al., 2018). Although the two levels of narrative organization are different theoretical concepts, they are related – and even interwoven – in children's oral narratives (Berman, 1988; Cain, 2003).

A methodological and much-used means for understanding the interrelations of a complex set of variables is offered by structural equation modelling, which allows for the modelling and testing of complex patterns by means of latent variables that represent theoretical constructs such as language skills. Most studies that examine the relationship between dimensions of oral language longitudinally using latent variables have focused on the core components (e.g. vocabulary and grammar) without including narrative skills. However, if we take a closer look at the studies that in fact have included narrative skills, their interrelations appear quite mixed. A study by the LARRC (2015) used latent variables in specifically addressing the relationship between narrative skills and core language skills in children at different ages. The study examined oral language skills (vocabulary, grammar and narrative skills) in a cross-sectional study of children aged 4–8. Narrative skills were measured with both expressive and receptive measures, and the assessment involved judging the inconsistency of stories, the picture arrangement of story structures and sentence arrangement in terms of building a consistent story. The study showed that whether the oral language skills included narrative skills or not varied across age. In pre-kindergarten the results were not conclusive, but in kindergarten oral language was best represented as a unitary skill that included narrative skills. However, when the children

reached first and second grades, core language skills appeared more differentiated from narrative skills. In third grade, oral language was further differentiated into three related dimensions (vocabulary, grammar and narrative skills; LARRC, 2015).

Some studies have used latent variables with narrative skills integrated in an oral language construct. For instance, Storch and Whitehurst (2002) found that the best-fitting model was one in which a narrative retelling (the Bus Story test) was part of a general oral language construct also consisting of vocabulary and grammar measures. However, it should be noted that this study did not set out to test the dimensionality of narrative skills versus other language skills. Considered together with other studies of dimensionality reviewed above, a trend of multidimensionality of language skills increasing with age is revealed. Notably, there appears to be support for the idea that narrative skills and core language skills (e.g. vocabulary and grammar) are separate constructs, but this seems mainly to be the case in older samples (LARRC, 2015; Tomblin & Zhang, 2006).

The predictive relationship between narrative retell and core language skills

Interest in narrative production as a way of assessing children's language skills increased significantly in the late 1980s when Bishop and Edmundson (1987) published their influential study demonstrating the high predictive value of narrative skills using the Bus Story Test. The authors examined the extent to which the results of a broad battery of language tests in 87 language-impaired children at age 4 predicted language group outcomes (resolved or persistent language impairments) at ages 4.5 and 5.5. Discriminant analysis showed that narrative skills comprised the single strongest predictor and accounted for as much as 83% of the group outcome. In addition, the authors found a positive relationship between low narrative skill scores and poor outcomes (persistent language impairments) and concluded that 'a child who at 4 years of age is unable to give even a simplified account of a sequence of events in a story accompanied by pictures is likely to have a poor outcome' (p. 169). This suggests that narrative skills may be a stronger predictor of language outcomes in children who perform poorly than in children who do well on narrative retell. Specifically, it is possible that narrative retell at age 4 is particularly predictive of later language skills in poor narrators; measures of narrative skills might capture underlying skills that are particularly significant for language development in the poorest performers. The predictive relation between levels of mastery of narrative retell and later language skills is unclear. This will be examined in this study.

Stothard et al. (1998) followed up with the children in the Bishop and Edmundson sample ($N=71$) and found that poor language outcomes at the age of 5.5 years were related to language and literacy outcomes at age 15. Similarly, Paul and Smith (1993) examined the language development of children from ages 2 to 5 ($N=55$) and found that narrative retell as measured by the Bus Story Test differed between late bloomers and children with expressive language delay at age 4. Finally, Botting et al. (2006) examined language skills (vocabulary and grammar), narrative retell (the Bus Story) and reading in a sample of approximately 200 children with a history of specific language impairment at the ages of 7 and 11. The authors found moderate correlations between Bus Story scores at age 7 and measures of vocabulary and grammar at ages 7 ($r=0.35-0.53$) and 11 ($r=0.43-0.47$). The sample was recruited from specialist language placements; it is

uncertain whether the strength of the relationship between narrative skills and later language holds in an unselected cohort. This is a focus of this study.

Approach and purpose

We aim to expand the existing knowledge base regarding the relationship between narrative retell, core language and listening comprehension skills in an unselected cohort of children followed from ages 4 to 7. Inspired by previous studies (e.g. Bishop & Edmundson, 1987; Heilmann et al., 2010), we chose narrative retell as a measure of narrative skills to encourage the 4-year-old children in our sample to produce more complex narratives (Heilmann et al., 2010) with more syntactically complex utterances (Mäkinen et al., 2018). Narrative retell appears to be a rather reliable way of measuring narrative skills for young children who might not have the experience, knowledge or cognitive maturity to tell a self-generated story at demand (e.g. Berman, 1988).

Although some attempts have been made to identify measurement errors (e.g. Vandewalle et al., 2012), the abovementioned studies used manifest variables for which measurement errors represent threats to the validity of the findings (Cole & Preacher, 2014). We applied error-reduction strategies, such as the use of latent variables, to reduce the effects of measurement error, as recommended by Cole and Preacher (2014), and based our analysis on the total variation of scores in the sample. The following questions guided the study:

1. How do narrative skills, as measured with the Bus Story Test, relate to a set of core language skills (vocabulary, grammar and verbal memory) at age 4?
2. How do the results relating to narrative skills predict core language skills (vocabulary, grammar and verbal memory) and listening comprehension at age 7?
3. Do narrative skills have unique, predictive value for core language skills (vocabulary, grammar and verbal memory) or listening comprehension at age 7 after controlling for core language skills (vocabulary, grammar and verbal memory) at age 4?
4. Are there stronger relationships between narrative skills at age 4 and core language skills (vocabulary, grammar and verbal memory) and listening comprehension at age 7 for children with poorer narrative skills at age 4?

Answering these questions has both theoretical and practical implications. Finding that narrative skills go together with the set of core language skills (vocabulary, grammar and verbal memory) at age 4 will support the hypothesis that language is a unidimensional construct. However, if we find support for the opposite hypothesis, that narrative skills are a separate ability from core language skills, it could suggest that some aspects of language are more differentiated in young children.

Finding that narrative skills predict the core language construct at later time points would provide support for the hypothesis that narrative retelling is important for the development of language skills. However, if retelling skills at age 4 do not predict the development of core language skills (vocabulary, grammar and verbal memory) or listening comprehension at age 7 beyond core language skills at age 4, then whether the assessment of narrative skills is necessary when predicting later core language skills would be questioned. It would then seem that what is captured can be largely captured through conventional language tests. Nevertheless, finding a stronger relationship

between narrative skills at age 4 and core language at age 7 in children with poor narrative skills at age 4 would indicate that the assessment of narrative skills is particularly important in children with poorer language skills.

Methods

Participants and procedures

The participants comprised 215 children (109 boys and 106 girls, age in years: $M=4.3$, $SD=0.19$) out of a cohort in an average Norwegian municipality. All parents gave consent for their children to participate in the study. All children attended Norwegian day-care centres. Children with intellectual disabilities or sensory impairment who were unable to complete the test battery and children who spoke Norwegian as a second language at the beginning of the study were excluded. The mean education level of the parents was close to the national average, as reported in a parent questionnaire (highest level of completed education). The children had average nonverbal IQ (a mean standard score of 10 according to the block design in Wechsler Preschool and Primary Scale of Intelligence [WPPSI]/Wechsler Intelligence Scale for Children [WISC] at ages 4 and 7 years). They were assessed individually in daycare centres and schools at age 4 and in second grade (age in years: $M=7.3$, $SD=0.18$) by trained research assistants with general clinical training and specific training in administering the assessment battery.

The attrition from the initial assessment at age 4 to the second assessment at age 7 was 14.9% ($n=32$). Attrition due to the exclusion of a group of children who were younger than the main group and therefore had started school a year later was 10%. The remaining attrition, 4.9%, was the result of children moving to other (remote) school districts.

Measures

Children were assessed using a Norwegian version of the Renfrew Bus Story Test at age 4. At both ages 4 and 7, children were additionally assessed using measures of receptive vocabulary, sentence memory and grammatical knowledge. At age 7, they were also assessed using a listening comprehension test.

The Renfrew Bus Story Test (Renfrew, 1997) is a retelling task in which a child is told a story about a 'naughty bus' while looking at 12 corresponding pictures of how the bus behaved. The scaffolding given using pictures and a model story (Heilmann et al., 2010) might be beneficial for young children's narrative production (see Berman, 1988) and appears to be effective based on the results of previous studies (Bishop & Edmundson, 1987; Mäkinen et al., 2018). Immediately afterwards, the child is asked to retell the story while looking at the pictures. In administering the test, only minimal neutral prompts were given ('Yes?', 'Can you tell me more?'). The narratives were audio-recorded and transcribed. An *information score* and the *mean length* of the five longest utterances (*length score*) were calculated manually and in accordance with the manual. A scoring form where the most frequent responses, including the actors, and key elements of the story were listed in the correct narrative order was used to increase the reliability of the scoring. The *information score* was calculated by assigning points for the number of relevant information units the child included in his or her retelling.

Information had to be given in relation to the corresponding pictures; thus, the order of the presented information had to follow the story structure and indicate the child's mastery of story organization (story coherence). Two points were awarded for *essential* information, and one point was awarded for *subsidiary* information. If the referent was not specified, one point was deleted (e.g. *The train drove into a tunnel* [2 p.]. *The bus ran into the street* [2 p.] versus *The train drove into a tunnel* [2 p.]. *He [the bus – not specified] ran into the street* [1 p.]). Consequently, the information score also implicitly takes story cohesion into account. The maximum score for story information was 52. The *length score* was calculated by counting the number of words in the five longest utterances and calculating the mean score. Repetitions, fillers and words such as *and* or *then* were not counted. This score provides information about the child's linguistic skills at the narrative micro level. Reliability was calculated using generalizability theory. Eighteen percent of the narratives (N=40) were double-coded. The index of dependability was $\phi = .980$ for the information score and $\phi = .885$ for the score on sentence length, indicating high reliability (Nordeide, 2008).

Receptive vocabulary. A Norwegian version of the British Picture Vocabulary Scale (BPVS; Dunn et al., 1997; Norwegian version: Lyster et al., 2010) was used to measure receptive vocabulary. In this test, the child is presented with a word and asked to select the corresponding picture/drawing from among four drawings. The maximum score was 144, and the stop criterion is eight errors or more out of a block with 12 items.

Verbal memory. The Sentence Memory subtest of the Norwegian language screening test 'Language 6–16' (Ottem & Frost, 2005) was used to assess verbal working memory. In this test, the child is asked to repeat sentences of increasing length and syntactic complexity. The original task consisted of 16 items. However, in this study five simpler items were added from the Norwegian version of Wechsler Preschool and Primary Scale of Intelligence–Revised (WPPSI-R; Wechsler, 2002), and the maximum score was therefore 21. The stop criterion was three incorrectly repeated sentences.

Grammatical knowledge. The Norwegian version of the Grammatical Closure subtest from the Illinois Test of Psycholinguistic Abilities (Kirk et al., 1968) was used to assess expressive grammatical knowledge. In this test, the child is presented with a series of pictures followed by incomplete spoken sentences and asked to complete the sentences. The maximum score was 33. The stop criterion was six consecutive incorrect items at age 4, whereas all items were presented at age 7.

Listening comprehension. A Norwegian version of the Neale Analysis of Reading Ability (NARA II; Neale, 1997) was used to measure listening comprehension in second grade (at age 7). In this test, the child is told a story by the test administrator and is then asked to respond orally to open-ended comprehension questions. To provide correct answers, the child has to draw inferences about the relationships in the stories. The test consists of six different stories of increasing length and complexity. The maximum score was 44. The stop criterion was four or more incorrect answers for one story.

Analyses

We used structural equation modelling with latent variables to answer the research questions. We first examined the factor structure of the measures using confirmatory factor analyses (CFAs) to assess how well different indicators fit the hypothesized constructs. Next, we used latent variable path models to assess how the Bus Story Test was related to the different language constructs over time, while an adjusted chi-square likelihood ratio test was used for model comparison (Bryant & Satorra, 2012). A major advantage of using latent variables for all constructs is that we can adjust for measurement error using only true-score variance to estimate the regressions and covariances between them (e.g. Little, 2013; Moosbrugger et al., 2009). Furthermore, we estimated interactions, such as whether the strength of the relationship between the Bus Story Test and later language skills varied with the Bus Story Test scores (specifically whether poorer Bus Story Test scores were associated with a stronger influence of narrative skills on later language skills). None of the studies reviewed the examined interactions using continuous variables in this manner.

Finally, all analyses using latent variables were performed with Mplus version 7.4 (Muthén & Muthén, 1998–2016). Full information maximum likelihood (FIML) was used to handle missing data under the ‘missing at random’ assumption, and all available information for each individual was used. We applied robust (Huber–White) standard errors for all estimated parameters and a scaled goodness-of-fit chi square for statistical inference. Model fit was evaluated based on commonly recommended goodness-of-fit indices (Hu & Bentler, 1999), including the chi-square test of exact model fit, the root mean square error of approximation (RMSEA: $\leq 0.08 = \text{acceptable}$, $\leq 0.05 = \text{good}$) to assess close fit, the comparative fit index (CFI: $\geq 0.90 = \text{acceptable}$, $\geq 0.95 = \text{good}$) contrasting to a null independence model and the standardized root mean square residual (SRMR: $\leq 0.08 = \text{good}$).

Results

Means, standard deviations, maximum scores and reliabilities for all measures at all time points, as well as correlations between the measures, are shown in Supplemental Tables 1 and 2. With a few exceptions, all measures had satisfactory alpha reliabilities, the test scores were normally distributed within the robust range (values between -2 and 1.20 for skewness and between -0.72 and 6.41 for kurtosis) and there were moderate to strong correlations between the different measures.

The relationship between narrative skills and a set of core language skills

Figure 1(a) shows the CFA model we estimated to test whether narrative skills, assessed with the two indicators from the Bus Story Test (information and sentence length), and the core language skills tests (vocabulary, grammar and verbal memory) at age 4 showed a better fit as a one-factor model than as a two-factor one. The model fit indices recommended by Hu and Bentler (1999) do not indicate a good fit to the data for the one-factor model at age 4: $\chi^2(5) = 45.00$, $p < 0.001$, RMSEA = 0.202 (90% confidence interval [CI] = 0.150–0.258), CFI = 0.872, TLI = 0.744 and SRMR = 0.092. The two-factor model, however, had an excellent fit to the data: $\chi^2(4) = 4.90$, $p = 0.30$, RMSEA = 0.034 (90%

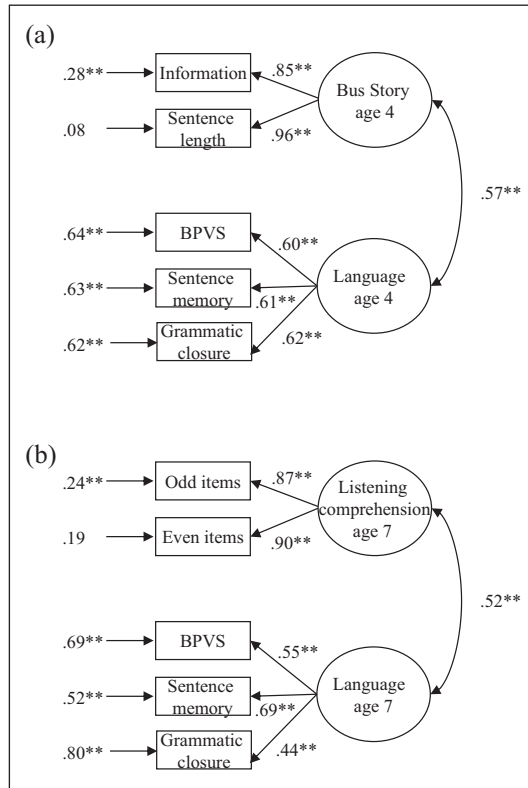


Figure 1. (a) Narrative Production and Language as Two Factors (Age 4): *Model Fit:* $\chi^2(4)=4.90$, $p=0.298$, RMSEA=0.034 (90% CI=0.000–0.118), CFI=0.997, TLI=0.993 and SRMR=0.021. (b) Listening Comprehension and Language as Two Factors (Age 7): *Model Fit:* $\chi^2(4)=2.51$, $p=0.643$ RMSEA=0.000 (90% CI=0.000–0.088), CFI=1.00, TLI=1.01 and SRMR=0.018.

CI=0.000–0.118), CFI=0.997, TLI=0.993 and SRMR=0.021. A chi-square difference test showed that the results supported a two-factor solution rather than a one-factor solution, $\Delta\chi^2(1)=35.29$ ($p < 0.001$), suggesting that the Bus Story Test and the core language skills tests represented two different constructs at this age. The correlation between the latent Bus Story factor and the language factor was 0.57.

Figure 1(b) shows the CFA we estimated for the language outcome variables at age 7. The results supported a two-factor model with core language skills and listening comprehension rather than a one-factor solution: $\Delta\chi^2(1)=56.02$, $p < 0.001$. Thus, at 7 years these two factors were best conceptualized as two different but correlated constructs: $r=0.52$, $\chi^2(4)=2.51$, $p=0.64$, RMSEA=0.000 (90% CI=0.000–0.088), CFI=1.00, TLI=1.01 and SRMR=0.018. The model fit for the one-factor solution was as follows: $\chi^2(5)=34.69$, $p < 0.001$, RMSEA=0.177 (90% CI=0.124–0.234), CFI=0.88 and TLI=0.77. The correlation between the core language factor and listening comprehension was 0.52.

Narrative skills at age 4 as a single predictor of core language skills and listening comprehension at age 7

Figure 2 shows the relationship between the Bus Story Test and the latent language factor at age 7 (second grade) when the Bus Story Test is the single predictor. Bus Story at age 4 explained 16.8% of the variance in listening comprehension or core language skills at age 7. This model had a good fit: $\chi^2(11)=7.66, p=0.74, RMSEA=0.000$ (90% CI=0.000–0.058), CFI=1.00, TLI=1.01 and SRMR=0.022.

The unique predictive value of narrative skills for core language skills and listening comprehension at age 7 after controlling for core language at age 4

When we included the language factor (vocabulary, grammar and verbal memory) at age 4, narrative skills did not explain additional variance beyond the 58% explained by the

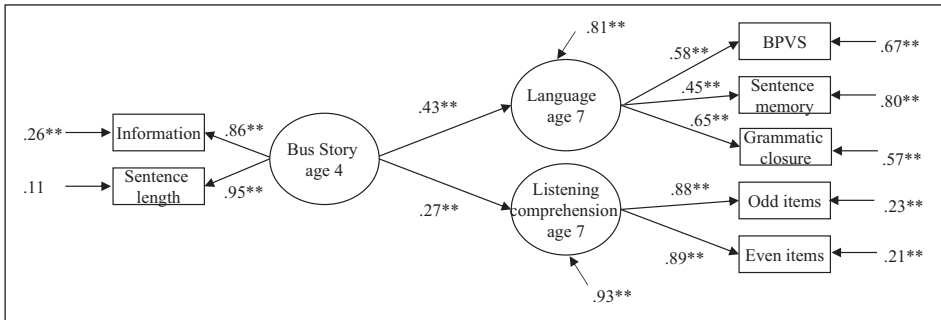


Figure 2. The Bus Story Test at Age 4 as a Single Predictor of Language and Listening Comprehension at Age 7. *Model Fit:* $\chi^2(11)=7.66, p=0.74, RMSEA=0.000$ (90% CI=0.000–0.058), CFI=1.00, TLI=1.01 and SRMR=0.022.

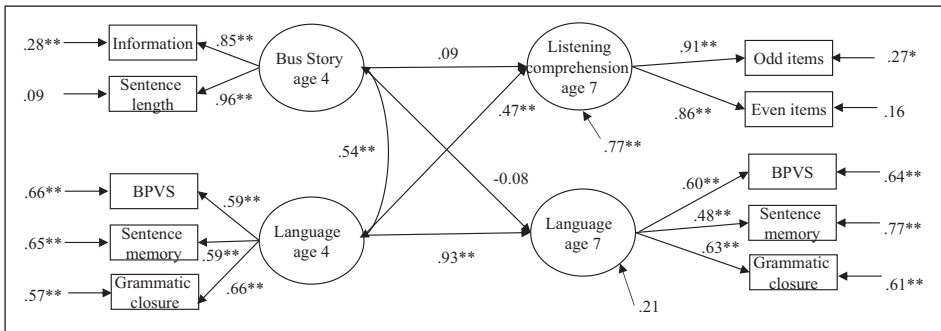


Figure 3. The Bus Story Test and Language at Age 4 as Predictors of Language and Listening Comprehension at Age 7. *Model Fit:* $\chi^2(26)=19.48, p=0.815, RMSEA=0.000$ (90% CI=0.000–0.046) CFI=1.000, TLI=1.02 and SRMR=0.028. The Residuals of the Same Indicators Are Correlated.

core language factor. The model shown in Figure 3 had an excellent fit to the data: $\chi^2(26)=19.48, p=0.82, RMSEA=0.000$ (90% CI=0.000–0.046), CFI=1.000, TLI=1.02 and SRMR=0.028.

The relationship between different levels of narrative skills at age 4 and core language skills and listening comprehension at age 7

To examine whether the strength of the relationship with later language skills was related to the levels of narrative skills at age 4 and core language and listening comprehension at age 7, we added an interaction term between narrative skills at age 4 and core language skills. The coefficients of the interaction terms were not significantly different from zero and added little to the model fit (for listening comprehension, unstandardized estimate=0.001, $p=0.87$; for core language skills, estimate=-0.001, $p=0.93$). Therefore, the interaction was not included in the final models in Figure 3. In addition, the results indicated that there was no need to examine possible curvilinear effects.

Discussion

This study revealed several findings of importance for our understanding of narrative retell skills in relation to other aspects of language and language development. The study also has some strengths that increase the validity of our results. First, we used the full variation of scores without using a cut-off, thereby decreasing the risks associated with cut-off studies (e.g. phenomena such as regression to the mean, leading children to cross the arbitrary cut-off and accidentally end up in the ‘resolved’ group). Second, we used latent variables, avoiding the risks associated with manifest variables, such as measurement error (Cole & Preacher, 2014). Finally, we had a larger sample than those in most previous studies examining the Bus Story Test (Bishop & Edmundson, 1987 [N=87]; Stothard et al., 1998 [N=71]; Pankratz et al., 2007 [N=32]; Paul & Smith, 1993 [N=55]), and the socioeconomic background of the sample was similar to the national average.

The relationship between narrative skills and core language skills

Regarding our first research question about the relationship between narrative retell and core language skills, we found support for the idea that narrative skills and core language skills are related, but best conceptualized as different skills. Previous research does not provide a clear answer to the question of dimensionality, and our results appear to be in contrast to most previous studies that support a general language construct in preschool and school-aged children (Bornstein & Putnick, 2012; Colledge et al., 2002; Justice et al., 2017; Kendeou et al., 2009; Protopapas et al., 2012; Tomblin & Zhang, 2006). In terms of the studies explicitly examining the relationship between narrative skills and core language-level skills, our findings are in accordance with the findings for school-aged children in the LARRC (2015) study.

However, our results are in contrast to the results for same-age children; in the LARRC study, language and discourse were both part of a unidimensional construct in kindergarten. Still, both the language and narrative skill constructs in the LARRC study were

broader and consisted of a wider range of indicators than the constructs in our study. In particular, in the LARRC study, the discourse construct consisted of various tasks related to narrative discourse, such as comprehension monitoring, narrative structure and the ability to draw inferences. Some of these measures may be more closely related to a general cognitive factor than the information and sentence scores of the Bus Story Test. In fact, it could be argued that the way the narratives are coded according to the test manual makes it more dependent on language skills.

One reason why studies of dimensionality have divergent findings could be the different methods used. When analysing data with latent variables, one uses what is common for the indicators in the analysis, and the variation that they do not have in common is partialled out. An advantage of this is that one controls for measurement error. However, methodological variation between the indicators is also something they do not have in common, and this is partialled out (Little, 2013). This means that if the tests have different formats, it can make the construct appear more different than it is in reality. For instance, if measures of the core language skills rely on pointing at pictures (such as the BPVS), this is a rather different testing format than a narrative retell task and can in itself make the constructs load on two different factors. In our case the correlation between narrative skills and core language skills was 0.57, and this was most likely too little overlap to be explained by methodological variation alone. However, for future studies to strictly test the dimensionality of language, the tests should have the same formats. If one has a measure of vocabulary that involves pointing at pictures, the grammar and narrative tasks should be as similar to this as possible. The main difference should be the skills the measure targets and not their formats.

Narrative skills at age 4 as a single predictor of core language skills and listening comprehension at age 7

Regarding our second research question concerning narrative skills and later language skills, we found associations between narrative retell at age 4 and a core language construct (vocabulary, grammar and verbal memory) at the same age. Furthermore, there was also a relationship between age 4 narrative skills and age 7 listening comprehension and core language skills (vocabulary, grammar and verbal memory). These associations were moderate for core language skills where narrative skills explained 19% of the variation at 7 years and small for listening comprehension where narrative skills explained 7%.

Notably, listening comprehension and core language were best described as two different but related constructs, which contradicts recent papers by Justice et al. (2017) and Lonigan and Milburn (2017). The reason for the contradictions could be that our listening comprehension task had a different format than the tasks measuring core language. Other studies have a larger variation in their listening comprehension task than ours did. For instance, the model by Justice et al. (2017) included a test of inference making in the listening comprehension construct in addition to a traditional listening comprehension task.

We were surprised to find that the relationship between previous narrative retells correlated more strongly with core language skills than with listening comprehension. However, this was probably caused by the different formats of the tasks: the scoring of

Bus Story, focusing on smaller information units and grammatical complexity, might make it more similar to the core language tasks than the listening comprehension task. More studies should be carried out, which are designed to untangle the relationship between different discourse-level modalities, and between discourse-level and core language skills at different ages.

The explanatory value of narrative retell for later language skills might seem disappointing. However, it is important to note that the time span here is as long as 3 years in a period when children undergo very rapid development in language skills. It is therefore perhaps not surprising that a different language measure 3 years earlier does not predict a larger portion of the variation.

The unique predictive value of narrative skills for core language skills and listening comprehension at age 7 after controlling for core language at age 4

Regarding our third research question, we did not find any support for the idea that narrative skills predict variation in core language skills or listening comprehension at age 7 over and above earlier core language skills (i.e. the autoregressor). This finding appears to conflict with the findings of most previous studies that found strong relationships between narrative proficiency and later language skills (e.g. Bishop & Edmundson, 1987; Paul & Smith, 1993; Stothard et al., 1998). One reason for these contradictory findings could be that not all previous studies controlled for autoregressors or that they conducted analyses that are not comparable to this one. However, it should be noted that because the children's rank order for core language skills is highly stable across time, much of the variation is already explained by the autoregressor, making it difficult for another variable to explain additional variation.

An important question concerns the implications of the finding that narrative skills are strongly related to core language skills at the same time point but do not predict additional variation at 7 years when core language skills at 4 years are taken into account. A conservative interpretation of this finding could be that we should rather assess core skills such as grammar and vocabulary, as narrative skills do not predict anything beyond this or add information in addition to this. Still, the fact that narrative skills are highly related to core language skills at the same time point does offer some possibilities for assessment. A potential advantage of assessing narrative skills is that it could give more insight into how a child is actually able to use language in situations closer to real life than receptive tests in multiple choice formats (such as BPVS) or word definitions. In some sense, narrative skills can be viewed as the product of core language skills in combination with pragmatic skills, an index of how a child is able to use core language skills in daily tasks and communication. Furthermore, testing all core language skills with bespoke measures is time-consuming and could be highly demanding for a child. Narrative skills are less tedious, take less time and could perhaps be seen as a first index to determine whether it could be important to further assess core skills. Therefore, even though their predictive value is limited, examining narrative skills might still be of clinical value.

Notably, in the context of interventions a recent systematic review showed that measures of narrative skills had a larger mean effect in oral language interventions than measures of vocabulary or grammar (Rogde et al., 2019). More precisely, vocabulary and grammar had mean effects of 0.17 standard deviation units (d), while narrative measures showed an effect of 0.42 standard deviation units. It could be speculated that the multifaceted nature of narrative skills could perhaps make them more sensitive to capture intervention effects than other kinds of language measures. The reason for this could be that narrative skills more closely align with the skills that are typically targeted in oral language interventions, such as joint book reading, constructing short stories and expressive language, than the traditional measures of core language skills. It might also be a question of skill constraints (Paradis, 2005): the skills needed to produce a coherent narrative at the macro level can be argued to be constrained to a limited set of knowledge (e.g. macro-structure; Stein & Glenn, 1979) compared with other skills such as vocabulary, in which knowledge is unconstrained. However, from a research perspective and particularly for intervention studies it could be valuable to assess narrative skills to gain more insight into the nature of improvements due to interventions.

The relationship between different levels of narrative skills at age 4 and core language skills and listening comprehension at age 7

Regarding our final research question, we did not find differential predictive validity for narrative skills at age 4 for listening comprehension or core language at age 7 between poorer performers and stronger performers. Previous studies using cut-off criteria concluded that narrative skills (measured with the Bus Story Test) comprise a particularly strong predictor of whether a child belongs to a language-impaired group or a non-language-impaired group (Bishop & Edmundson, 1987; Pankratz et al., 2007; Stothard et al., 1998). In particular, Bishop and Edmundson (1987) found a stronger relationship between the lowest range of scores for narrative skills and the risk of ending up in a poor-outcome group, thus inspiring us to examine interaction effects.

Our sample consisted of unselected children, but the prevalence of language disorders of unknown origin is approximately 7% (Norbury et al., 2016); therefore, one would presume that some of the children have language impairments that were not yet identified at age 4. Nevertheless, the interaction analyses did not support a stronger relationship between narrative skills at age 4 and core language or listening comprehension at age 7 in children with low scores for the narrative assessment than in children with high scores. Therefore, our findings are in contrast to those of Bishop and Edmundson (1987) and of other studies including children with language disorders (e.g. Norbury et al., 2014). However, it should be noted that our study had restricted power to detect whether predictive strength was dependent on very low levels of scores for narrative skills.

Final remarks and implications for future studies

Task effects in the assessment of oral narratives should be mentioned, as there are several studies that have examined the sensitivity of different procedures. For example, syntactic

complexity appears to increase in retellings compared with self-generated stories (Mäkinen et al., 2018; Westerveld & Moran, 2013), and the use of prompts (verbal, picture) appears to improve coherence (Cain, 2003). Consequently, different narrative task procedures may yield different results. It is important to emphasize, however, that previous studies have found moderate-to-strong correlations between different aspects of retelling and story generation (Mäkinen et al., 2018; Westerveld & Moran, 2013). Although retell provides a model story for the children and assist them in understanding the story structure (Heilmann et al., 2010), these correlations suggest that narrative retell taps into narrative proficiency. Assessing the oral language skills of young children can be challenging; a retell task with picture support might be more motivating and relevant than many other test formats.

In terms of future studies, this study has some caveats that future studies could address. First, in contrast to the abovementioned studies, we did not aim to identify children with language impairments. Therefore, the results might not apply to clinical settings. Although some previous studies indicated similar patterns of dimensionality in children with and without language disorders (Lonigan & Milburn, 2017; Tomblin & Zhang, 2006), more studies are needed to determine whether the same pattern applies to various types of samples. Second, we did not examine the longitudinal relationship between narrative retelling at different time points. This could have added to our understanding of the development of narrative retell in relation to core language skills and, therefore, should be a focus in future studies. Third, when assessing narrative skills, we used a particular retelling task with a particular system for scoring the narratives. Future studies should also include other narrative tasks (e.g. self-generated narratives) and ways of scoring the narratives that explicitly take narrative macro-structure into account and further examine this in relation to other discourse-level skills, as well as core language.

The findings of this study show that narrative skills are likely to be separate but partly overlapping with core language skills and that narrative skills in themselves are highly predictive of later core language skills, but that they do not add any explanatory value beyond core language skills at the previous point in time. Furthermore, the results do not support the idea that narrative skills are particularly sensitive to capture poor language skills, but more studies designed to answer this question are needed in order to settle this. Therefore, our results suggest that it can be useful to assess narrative skills in young children, but that in order to determine language problems such assessments should include measures of core language skills.

Authors contribution

JK, HNH, BEH and MML conceived the presented idea and all authors contributed in developing the theoretical models underlying the analyses. BEH and MML was main responsible for the study. JK and HNH participated in the data collection and data management. JK wrote the first draft of the introduction, method and discussion with support from MML and HNH. JB conducted the SEM-analyses and MML wrote the results section. All authors discussed the results, commented on and contributed to the final version of the manuscript. JK was in charge of the revisions, but all authors contributed significantly and approved the final version of the manuscript.

ORCID iD

Monica Melby-Lervåg  <https://orcid.org/0000-0001-6461-1167>

Supplemental material

Supplemental material for this article is available online.

References

- Berman, R. A. (1988). On the ability to relate events in narrative. *Discourse Processes, 11*(4), 469–497. <https://doi.org/10.1080/01638538809544714>
- Bishop, D. V. M., & Edmundson, A. (1987). Language impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52*, 156–173. <https://pubs.asha.org/doi/10.1044/jshd.5202.156>
- Bornstein, M. H., & Putnick, D. L. (2012). Stability of language in childhood: A multiage, multi-domain, multimeasure, and multisource study. *Developmental Psychology, 48*(2), 477–491. <https://doi.org/10.1037/a0025889>
- Botting, N., Simkin, Z., & Conti-Ramsden, G. (2006). Associated reading skills in children with a history of specific language impairment (SLI). *Reading and Writing, 19*, 77–98. <https://doi.org/10.1007/s11145-005-4322-4>
- Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(3), 372–398. <https://doi.org/10.1080/10705511.2012.687671>
- Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology, 21*(3), 335–351. <https://doi.org/10.1348/026151003322277739>
- Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19*(2), 300–315. <https://doi.org/10.1037/a0033805>
- Colledge, E., Bishop, D. V., Koeppen-Schomerus, G., Price, T. S., Happé, F. G., Eley, T. C., . . . Plomin, R. (2002). The structure of language abilities at 4 years: A twin study. *Developmental Psychology, 38*(5), 749–757. <https://doi.org/10.1037/0012-1649.38.5.749>
- Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale II*. National Foundation for Educational Research-Nelson.
- Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*, 154–166. [https://doi.org/10.1044/1058-0360\(2009/08-0024\)](https://doi.org/10.1044/1058-0360(2009/08-0024))
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Justice, L. M., Lomax, R., O'Connell, A., Pentimonti, J., Petrill, S. A., & Piasta, S. B., . . . Language and Reading Research Consortium (LARRC). (2017). Oral language and listening comprehension: Same or different constructs? *Journal of Speech, Language, and Hearing Research, 60*(5), 1273–1284. https://doi.org/10.1044/2017_JSLHR-L-16-0039
- Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contribution of oral language and decoding skills. *Journal of Educational Psychology, 101*(4), 765–778. <https://doi.org/10.1037/a0015956>

- Kim, Y.-S. G., Park, C., & Park, Y. (2015). Dimensions of discourse-level oral language skills and their relations to reading comprehension and written composition: An exploratory study. *Reading and Writing, 28*, 633–654. <https://doi.org/10.1007/s11145-015-9542-7>
- Kirk, S., McCarthy, J. J., & Kirk, W. D. (1968). *The Illinois test of psycholinguistic abilities*. University of Illinois Press.
- Language and Reading Research Consortium. (2015). The dimensionality of language ability in young children. *Child Development, 86*(6), 1948–1965. <https://doi.org/10.1111/cdev.12450>
- Little, T. D. (2013). *Longitudinal structural equation modelling*. Guilford Press.
- Lonigan, C. J., & Milburn, T. F. (2017). Identifying the dimensionality of oral language skills of children with typical development in preschool through fifth grade. *Journal of Speech, Language, and Hearing Research, 60*, 2185–2198. https://doi.org/10.1044/2017_JSLHR-L-15-0402
- Lyster, S.-A. H., Horn, E., & Rygvold, A.-L. (2010). Ordforråd og ordforrådsutvikling hos norske barn og unge. Resultater fra en utprøving av British Picture Vocabulary Scale I [Vocabulary and vocabulary development. Results from the Norwegian version of The British Picture Vocabulary Scale], Second Edition (BPVS II). *Spesialpedagogikk (Norwegian Journal of Special Needs Education), 9*(75), 34–44.
- Mäkinen, L., Soile, L., Ilaria, G., & Sari, K. (2018). Are story retelling and story generation connected to reading skills? Evidence from Finnish. *Child Language Teaching and Therapy, 34*(2), 129–139. <https://doi.org/10.1177/0265659018780960>
- Moosbrugger, H., Schermelleh-Engel, K., Kelava, A., & Klein, A. G. (2009). Testing multiple nonlinear effects in structural equation modelling: A comparison of alternative estimation approaches. In T. Teo, & M. S. Khine (Eds.), *Structural equation modelling in educational research: Concepts and applications* (pp. 10–30). Sense Publishers.
- Muthén, B., & Muthén, L. (1998–2016). *Mplus user's guide* (6th ed.).
- Neale, M. D. (1997). *The Neale analysis of reading ability* (Revised ed.). InferNelson.
- Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language, 41*(3), 485–510. <https://doi.org/10.1017/S030500091300007X>
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry, 57*, 1247–1257. <https://doi.org/10.1111/jcpp.12573>
- Nordeide, I. (2008). Reliabilitet ved bruk av Renfrew Bus Story test [Reliability using the Renfrew Bus Story Test] [Master's thesis]. The University of Oslo.
- Ottem, E., & Frost, J. (2005). Språk 6-16 screeningtest av språkvansker [Language 6-16 Screening Test] (Version 3) Bredtvet kompetansesenter. Oslo.
- Pankratz, M. E., Plante, E., Vance, R., & Insalaco, D. M. (2007). The diagnostic and predictive validity of the Renfrew Bus Story. *Language, Speech, and Hearing Services in Schools, 38*, 390–399. [https://doi.org/10.1044/0161-1461\(2007\)040](https://doi.org/10.1044/0161-1461(2007)040)
- Paradis, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184–202. <https://doi.org/10.1598/RRQ.40.2.3>
- Paul, R., & Smith, R. L. (1993). Narrative skills in 4-year-olds with normal, impaired and late-developing language. *Journal of Speech and Hearing Research, 36*, 592–598. <https://doi.org/10.1044/jshr.3603.592>
- Protopapas, A., Simos, P. G., Sideridis, G. D., & Mouzaki, A. (2012). The components of the Simple view of reading: A confirmatory factor analysis. *Reading Psychology, 33*, 217–240. <https://doi.org/10.1080/02702711.2010.507626>
- Renfrew, C. (1997). *Bus story test* (4th ed.). Winslow Press.

- Rogde, K., Hagen, Å. M., Melby-Lervåg, M., Lervåg, A., Hagen, Å. M., & Melby-Lervåg, M. (2019). The effect of language comprehension training on standardized tests: A systematic review. *Campbell Systematic Reviews*, 15(4), 1–37. <https://doi.org/10.1002/cl2.1059>
- Stein, N. L., & Glenn, C. G. (1979). *An analysis of story comprehension in elementary children (Vol. 2)*. Ablex.
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38(6), 934–947. <https://doi.org/10.1037/0012-1649.38.6.934>
- Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired pre-schoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research*, 41(2), 407–418. <https://doi.org/10.1044/jslhr.4102.407>
- Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence: The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cognitive Development*, 47, 82–95. <https://doi.org/10.1016/j.cogdev.2018.04.005>
- Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*, 49(6), 1193–1208. [https://doi.org/10.1044/1092-4388\(2006\)086](https://doi.org/10.1044/1092-4388(2006)086)
- Vandewalle, E., Boets, B., Boons, T., Ghesquière, P., & Zink, I. (2012). Oral language and narrative skills in children with specific language impairment with and without literacy delay: A three-year longitudinal study. *Research in Developmental Disabilities*, 33, 1857–1870. <https://doi.org/10.1016/j.ridd.2012.05.004>
- Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence (WPPSI-R). Manual (Norwegian Version 2008)*. Harcourt Assessment.
- Westerveld, M. F., & Moran, C. A. (2013). Spoken expository discourse of children and adolescents: Retelling versus generation. *Clinical Linguistics & Phonetics*, 27(9), 720–734. <https://doi.org/10.3109/02699206.2013.802016>