The Role of Inference Making and Other Language Skills in the Development of Narrative Listening Comprehension in 4–6-Year-Old Children

Janne Lepola  
*University of Turku, Rauma, Finland*

Julie Lynch  
*Saginaw Valley State University, University Center, Michigan, USA*

Eero Laakkonen  
*University of Turku, Finland*

Maarit Silvén  
*University of Turku, Finland*

Pekka Niemi  
*University of Turku, Finland*

**Abstract**

In this two-year longitudinal study, we sought to examine the developmental relationships among early narrative listening comprehension and language skills (i.e., vocabulary knowledge, sentence memory, and phonological awareness) and the roles of these factors in predicting narrative listening comprehension at the age of 6 years. We also sought to examine the role of inference-making skills as longitudinal and concurrent predictors of other language skills and listening comprehension from the age of 4 to 6 years. One hundred thirty Finnish-speaking children participated in the study. A theoretical model of the developmental relationships among the variables was proposed and the associations were analyzed by means of path analysis. Results showed that inference skills, assessed through picture-book viewing, made a significant and unique contribution to variation in later narrative listening comprehension. Inference skills also played an indirect role in narrative listening comprehension by making a significant contribution to vocabulary knowledge even after controlling for earlier vocabulary knowledge and sentence memory. Although vocabulary knowledge and sentence memory were related to concurrent narrative listening comprehension, they did not predict later listening comprehension over and above the autoregressor. The results are discussed in terms of the predictive validity and diagnostic sensitivity of inference skills assessments in listening comprehension. Implications for research and theory are also discussed.

Narratives surround children from their earliest language experiences. Young children experience narrative stories through shared book reading at home or day care, watching television programs with a narrative structure, and participation in talk about daily events. Thus, narrative listening
comprehension skills are developing well before a child is able to read and comprehend texts independently. According to the multicomponent model, as a child attempts to construct a coherent, meaning-based representation of a narrative, a number of cognitive processes are at work at the word, sentence, and text levels (Oakhill & Cain, 2007; Perfetti, Landi, & Oakhill, 2005).

At the word and sentence levels, children store phonological information in short-term memory to derive the meaning of words and integrate individual meanings into semantically more complex meanings (Gathercole & Baddeley, 1989). Thus, children need to draw on their vocabulary knowledge and verbal memory. At the text level, inference making is needed to identify characters and their motives, follow the plot of the story, and understand the explicit and implicit information in the story (Florit, Roch, & Levorato, 2011; Graesser, Singer, & Trabasso, 1994; Oakhill & Cain, 2007; van den Broek et al., 2005). Inference making refers to skills in identifying information that is left implicit in the text and integrating information across various parts of the text, as well as between the text and the listener’s background knowledge (Kendeou, Bohn-Gettler, White, & van den Broek, 2008; Paris & Paris, 2003).

Studying the development of children’s narrative listening comprehension is important both theoretically and practically. Theoretically, this area of research provides insights into the development of a very complex cognitive process. We still have much to learn about the development of the individual skills necessary for narrative listening comprehension (e.g., inference making, vocabulary), how these skills may influence each other across time, and how they become integrated to produce skilled listening comprehension. Practically, narrative listening comprehension is related to school achievement (Feagans & Appelbaum, 1986), and it is an important precursor to reading comprehension, as a very similar set of cognitive skills are needed to derive the meaning of aural and written stories (e.g., Kendeou, van den Broek, White, & Lynch, 2009; McKeown & Beck, 2006). Many children’s books and much instructional discourse in the primary grades are narrative by nature. Thus, a better understanding of the development of narrative listening comprehension may lead to more effective strategies to foster this ability in the preschool years and, presumably, promote later reading comprehension.

The first purpose of the present study was to analyze the developmental associations between narrative listening comprehension and more foundational language skills, such as vocabulary knowledge, sentence memory, and phonological awareness, and the roles of these factors in the prediction of narrative listening comprehension at age 6. The second purpose was to examine the role of inference-making skills assessed at 4 and 5 years of age in predicting narrative listening comprehension at age 6, for children who are fluent in speech production but are not yet readers (in the Finnish context). We were particularly interested in examining the developmental associations between text-level skills (i.e., early narrative listening comprehension, inference skills) and foundational language skills, as well as the direct and indirect contributions of these skills in the prediction of subsequent listening comprehension.

Despite the importance of inference skills in reading comprehension (Cain, Oakhill, & Bryant, 2004; Oakhill & Cain, 2011), the research on listening comprehension is scant (Kendeou et al., 2008), and longitudinal studies are even more rare. Although studies have suggested that children start to understand complex narratives by 4 years of age (Bruner, 1990; Kendeou et al., 2008; Skarakis-Doyle & Dempsey, 2008), there is a gap in our knowledge about the developmental dynamics among foundational language skills and inference-making skills and the roles of these factors in the development of narrative listening comprehension.

**Review of Research**

Although the unique roles of inference and foundational skills have been noted in research on listening comprehension (Florit et al., 2011; Kendeou et al., 2008), previous studies have been cross-sectional or have not examined whether inference skills predict the development of listening comprehension over and above the contribution of the autoregressor—that is, the contribution of previous performance on the same skill. The longitudinal design of the present study enables us to take into account the autoregressive effects and thus provide a clearer picture of the developmental relations between text-level and foundational language skills.

**Foundational Skills: Vocabulary, Sentence Memory, and Phonological Awareness as Predictors of Narrative Listening Comprehension**

The rate of learning meanings for words increases rapidly during the second and third years of life, and productive vocabulary size during this time is developmentally related to grammatical and literacy skills (Silvén, Poskiparta, Niemi, & Voeten, 2007; Snow, Burns, & Griffin, 1998). In fact, vocabulary knowledge has been shown to be one of the best predictors of narrative listening comprehension (Kendeou et al., 2008; Sénéchal, Ouellette, & Rodney, 2006) and reading comprehension (Muter, Hulme, Snowling, & Stevenson, 2004; Roth, Speece, & Cooper, 2002).

A longitudinal study by Sénéchal et al. (2006) showed that among children with an upper middle
class background, receptive vocabulary at kindergarten explained 7% and 8% of unique variance in listening comprehension at kindergarten and at grade 1, respectively, after controlling for parent literacy, child age, literacy skills, and phonological awareness. In a cross-sectional study with typically developing children, Florit, Roch, Altoé, and Levorato (2009) reported moderate correlations between the performance of a receptive vocabulary task and listening comprehension at the age of 4 ($r = .45$) and 5 years ($r = .40$). In addition, the results of a longitudinal study by Kendeou et al. (2008) showed that receptive vocabulary correlated 0.41 and 0.48 with narrative listening comprehension among the 4- and 6-year-olds, respectively. Moreover, Kendeou et al. showed that among the sample of 4-year-olds with middle class backgrounds, receptive vocabulary and inference skills were the two unique predictors of narrative listening comprehension.

In the present study, the development of children’s vocabulary knowledge was assessed by an age-appropriate word definition task. Children’s skill in defining words has been associated with the concurrent semantic and phonological skills among 4-year-olds (Silvén & Rubinov, 2010), as well as with listening comprehension ($r = .53$) among second and third graders (Wise, Sevcik, Morris, Lovett, & Wolf, 2007). What is more, recent studies of grade 3 (Tannenbaum, Torgesen, & Wagner, 2006) and grade 4 students (Ouellette, 2006) have shown that the association between word definition and reading comprehension is stronger than the association between receptive vocabulary and reading comprehension.

Understanding narratives may not only challenge young children’s access to word meanings (Ouellette, 2006) but also their verbal memory. When listening to a narrative, a child draws on verbal memory resources to derive the meaning of a word, temporarily hold that meaning, and integrate the meanings of several words in a sentence. Verbal memory resources are also necessary as the child integrates the meaning of the currently heard information to previous parts of the narrative and to his or her background knowledge (van den Broek & Lorch, 1993).

A number of studies have shown that two memory functions, verbal short-term memory (i.e., the capacity to store material over a short period of time) and verbal working memory (i.e., the capacity to maintain and manipulate information), play an important role in concurrent listening comprehension among preschool- and school-age children (Daneman & Blennerhasset, 1984; Dufva, Niemi, & Voeten, 2001; Florit et al., 2009). The contribution of these memory functions to listening comprehension was analyzed in a study by Florit et al. Their results suggested that among preschool-age children, both short-term memory and working memory were significant concurrent predictors of listening comprehension over and above receptive vocabulary and verbal IQ. Daneman and Blennerhasset showed that 3–5-year-old children’s listening comprehension (i.e., integrating the information across sentences) was strongly related to memory resources evaluated by word span ($r = .57$) and listening span ($r = .75$) tasks. Because of the complexity of the listening span tasks for young children (Daneman & Carpenter, 1980), a sentence repetition task was used in the present study. In other studies, researchers have also assessed verbal memory by asking children to repeat increasingly longer sets of sentences (Alloway, Gathercole, Willis, & Adams, 2004; Dufva et al., 2001).

Another foundational skill in the development of reading is phonological awareness. Phonological awareness is strongly related to the acquisition of word-decoding skills, as it allows children to segment the sounds of words and then master letter–sound correspondences (e.g., Ehri et al., 2001; Perfetti, 1994). The connection between phonological awareness and narrative listening comprehension is less clear. Phonological awareness in kindergarten-age children has been shown in some studies to be associated with both concurrent and subsequent listening comprehension (Dufva et al., 2001; Sénéchal et al., 2006). Phonological awareness appears not to contribute to later listening comprehension after the autoregressive effect of previous listening comprehension has been controlled for (Dufva et al., 2001). However, one study suggested that training phonological awareness improves listening comprehension in grade 1 (Poskiparta, Niemi, & Vauras, 1999, Table 4). A further reason to include phonological awareness in the current study is that it may indirectly contribute to listening comprehension through other foundational skills.

**Text-Levels Skills: Inference Making as a Predictor of Narrative Listening Comprehension**

Theoretical models of text comprehension (Gernsbacher, 1990; Kintsch, 1998; Perfetti, 1999) assume that inference making not only facilitates the formation of a situation, or mental, model of the text but also strengthens memory for the text. Thus, information that is not integrated into a mental model is more fragile and prone to be forgotten (Kintsch, Welsch, Schamolhofer, & Zimny, 1990). In line with this notion, inference skills have been found to contribute to young children’s ability to memorize and understand literal as well as inferred meanings of stories that they have listened to, watched, or both (Florit et al., 2011; Kendeou et al., 2008).

Longitudinal and intervention studies have indicated that a child’s ability to draw inferences is causally related to the development of text comprehension.
The longitudinal study by Kendeou et al. showed that children’s inference skills made a significant contribution to narrative listening comprehension both at the ages of 4 and 6 years over and above receptive vocabulary and phonological awareness. The strength of the relationship between inference skills and concurrent narrative comprehension was found to increase with age (Kendeou et al., 2008). Moreover, the results of a cross-sectional study by Florit et al. (2011) showed that after controlling for age, verbal memory, receptive vocabulary, and verbal IQ, inference-making skills still accounted for a small (2%) but yet significant amount of variance in narrative listening comprehension among preschoolers.

Paris and Paris (2003) showed that first- and second-grade students’ responses to inference-making questions, assessed through picture book viewing, were significantly correlated with reading comprehension scores derived from a recall of narrative elements and answering literal and inferential comprehension questions ($r = .28–.30$). The unique role of inference skills in comprehension development has been examined by Oakhill and Cain (2011). Their study showed that inference skills in children ages 8 or 9 years predicted unique variance in later reading comprehension (10–11 years) even when the contributions of earlier reading comprehension, receptive vocabulary, and verbal working memory were controlled for. The present study adds to this knowledge by asking whether inference-making skills are directly or indirectly related to the development of narrative listening comprehension even prior to reading acquisition.

In previous studies, the associations between inference skills and listening comprehension were assessed by various tasks. Children’s listening comprehension was evaluated in the Kendeou et al. (2008) study by computing the total number of recalled events in two stories and their centrality to the story structure as well as by answers to comprehension questions. The recalled events and answers to the questions that did not match the literal events in the narrative, but reflected story information and world knowledge, were coded as inferences. The inferential task in the study by Florit et al. (2011) consisted of short sentences referring to familiar events followed by knowledge-based and text-based questions, whereas listening comprehension was evaluated by answers to five explicit and five implicit comprehension questions about a story. Each question was followed by a multiple-choice task with four possible pictorial alternatives. Although the inference task in the study by Florit et al. was designed to take account of confounding factors such as memory resources, the researchers did not control for sentence and narrative comprehension. Different types of inferences were coded in the study by Kendeou et al.; however, both listening comprehension and inference-making scores were derived from the same recall of a narrative text, but no memory test was used.

In the present study, inference skills were assessed through implicit comprehension questions after picture book viewing, whereas narrative listening comprehension was evaluated through a retelling task and literal comprehension questions. According to Paris and Paris (2003), comprehension of pictorial narrative demands the same kind of skills as text-based stories, such as inference skills and knowledge of the narrative elements. Based on confirmatory factor analysis, Lepola, Peltonen, and Korpilahti (2009) showed that scores on these implicit comprehension questions measured inference skills among 4-year-olds. The inference skills overlapped with but were independent from a language factor reflecting vocabulary, verbal memory, and listening comprehension. Based on previous studies showing that children’s abilities to make inferences generalize across aural, televised, and written stories (Cain, Oakhill, & Bryant, 2004; Kendeou et al., 2008), we assumed that inferences assessed through picture book viewing were related to subsequent listening comprehension, either directly or indirectly via language skills.

Foundational Skills, Text-Level Skills, and the Development of Narrative Listening Comprehension

Regarding the developmental relationship between reading comprehension and inference skills, the study by Oakhill and Cain (2011) showed that in school years 3–6 in the United Kingdom, inference skills partially mediated the contribution of prior comprehension to subsequent comprehension. This is in line with studies that showed that 7- and 8-year-old children with poor reading comprehension are less likely to make inferences than children with good comprehension (Cain & Oakhill, 1999; Oakhill, 1984). Vocabulary and listening (or reading) comprehension have been shown to have a reciprocal relationship in children with specific reading comprehension difficulties (Cain & Oakhill, 2011), as well as in children from an unselected population (Stanovich, Cunningham, & Freeman, 1984; Verhoeven & Leeuwe, 2008). Thus, better listening comprehension leads to greater opportunities to learn vocabulary, and increased vocabulary size, in turn, seems to result in better listening comprehension. Based on a study by Sénéchal, Thomas, and Monker (1995), there is indirect evidence that 4-year-olds with larger vocabularies learn more words in storybook reading sessions than their age mates with less developed vocabularies. What is more, when shared book reading is characterized by active meaning-related talk (e.g., predicting, drawing inferences), it facilitates...
the growth of vocabulary (Ewers & Browson, 1999; Hindman, Connor, Jewkes, & Morrison, 2008).

Consequently, the relationship between narrative listening comprehension and vocabulary may be at least partially mediated by inference skills. Oakhill and Cain (2011) failed to find evidence for the unique role of comprehension in vocabulary growth, probably because they included a variety of comprehension-related skills such as inference making and comprehension monitoring. A study by Cain, Oakhill, and Lemmon (2004) showed that children with both poor vocabulary and poor reading comprehension were impaired at inferring the meanings of novel words from context. Thus, listening to a narrative does not ensure that a child is learning the meanings of novel words (Cain, Oakhill, & Lemmon, 2004). Rather, spontaneous inference making or instruction that explicitly teaches how to derive the meaning of unknown words from the text, such as looking for clue words, is needed (Nash & Snowling, 2006). Likewise, Paris and Paris (2003) showed that the scores for inference questions correlated significantly not only with subsequent reading comprehension (r = .35) but also to some extent with vocabulary skills (r = .28) assessed through the Gates–MacGinitie Reading Test. A caveat here is that these findings are mainly based on studies of readers.

Studies have also shown that verbal memory and vocabulary knowledge are developmentally related to each other in children between 3 and 7 years of age (Gathercole & Adams, 1993; Gathercole & Baddeley, 1993; Gathercole, Willis, Emslie, & Baddeley, 1992; Silvén & Rubinov, 2010). For instance, Gathercole and Baddeley (1989) reported that at age 4, verbal short-term memory was a significant predictor of vocabulary at age 5, even after the contribution of prior vocabulary and nonverbal IQ were taken into account. Although vocabulary knowledge contributes to short-term memory performance beyond age 5, vocabulary has not been found to uniquely predict short-term memory between the ages of 4 and 5 years (Gathercole et al., 1992).

Phonological awareness has been shown to be related to predictors of listening comprehension. For instance, Dufva et al. (2001) showed that phonological awareness at age 6 was related to concurrent verbal memory. Phonological awareness tasks tap verbal memory, as they demand the storage and processing of oral language. Also, as Crain and Shankweiler (1988) have suggested, individual differences in verbal memory may arise from difficulties in phonological processing. Thus, phonological awareness may indirectly contribute to listening comprehension through sentence memory.

**Present Study**

The aims of this two-year longitudinal study were, first, to examine whether foundational language skills, such as vocabulary, phonological awareness, and sentence memory, have a developmental relationship with narrative listening comprehension. Second, and more important, we ask whether inference-making skills uniquely explain listening comprehension after the foundational language skills have been taken into account. On the basis of previous research, we constructed a conceptual model of the relations among the observed variables (see Figure 1).

We built the final model in two phases by means of path analysis. The first model includes the developmental associations among narrative listening comprehension, vocabulary, sentence memory, and phonological awareness. The second, extended model involves inference skills as well as the above language skills that are related to the development of narrative listening comprehension. Thus, after having delineated the contributions of foundational skills to the development of listening comprehension, we studied whether there is room left for the direct or indirect effects of inference skills. Finally, we carried out a hierarchical regression analysis to examine whether concurrent inference skills contribute to narrative listening comprehension over and above the concurrent language skills.

On the basis of previous research, we hypothesized that both inference skills and vocabulary measured at 5 years of age (time 2) would contribute to individual differences in listening comprehension at 6 years of age (time 3). Likewise, we assumed that listening comprehension at time 2 would be predicted by time 1 inference skills and vocabulary, as well as by concurrent sentence memory (Dufva et al., 2001; Silvén & Rubinov, 2010). Of these three factors, we hypothesized that vocabulary would be the most influential (Florit et al., 2009; Oakhill & Cain, 2011). We also assumed that listening comprehension, inference skills, vocabulary, and sentence memory at time 2 would be predicted by the same skills measured at time 1.

Regarding the indirect paths to listening comprehension, we hypothesized that vocabulary knowledge at time 2 would mediate the relationships between time 1 sentence memory and inference skills and time 3 listening comprehension. We further assumed that inference skills would contribute to vocabulary knowledge (Hindman et al., 2008; Oakhill & Cain, 2011). Based on previous studies (Kendeou et al., 2008; Oakhill & Cain, 2011), we also assumed that listening comprehension at time 1 would have an indirect effect on listening comprehension at time 3 via inference skills at time 2. Thus, the better the children’s memory for the narrative text, the better they would be at making inferences, which in turn would predict subsequent listening comprehension. We also hypothesized that phonological awareness at time 1 would contribute to listening comprehension through sentence memory at time 2.
Method

Participants

This longitudinal study involved 130 typically developing, Finnish-speaking children (62 boys and 68 girls). All parents of children turning 4 years old during 2007 from 16 day-care centers (24 preschool groups) were asked for permission for their children to participate in the study. Only 4% of the parents refused to give written consent. The day-care centers represented socioeconomically varied districts from two towns with 176,000 and 14,500 inhabitants, respectively. Based on a questionnaire filled out by the mothers, their average educational level was slightly above that of the distribution reported for Finnish females between 25 and 49 years of age (Statistics Finland, 2007): 16% had a master’s degree, 50% had a BA or vocational college degree, 24% had vocational education or a high school diploma, and 10% had no vocational education. At the beginning of the study, the mean age of the children was 50.5 months (range = 45–56 months).

The participating 130 children were a subsample of the original sample of 149 children. The parental questionnaire and information from preschool teachers indicated no severe neurological or sensory difficulties. Parental reports indicated that four children suffer from extensive language impairment. These children were excluded to avoid outlier problems.

Four children had a home language other than Finnish and were therefore excluded. Seven families moved from the area, and their children were not assessed again. Two children refused to cooperate with the experimenter, and the assessment was terminated. In addition, the data of two children were not accessible because of a recording failure.

The Early Education System in Finland

The Finnish early education and care system differs to some extent from the systems in other European countries and the United States. Children enter kindergarten (the term preschool is used in Finland) in August of the year they turn 6, and formal reading instruction begins when a child enters grade 1 in August of the following year. All 6-year-olds are entitled to free kindergarten including free lunch. Practically all 6-year-olds participate in kindergarten, which comprises 700 hours per year. More than 70% of 4- and 5-year-olds attend municipal day care provided at day-care

Figure 1. Conceptual Model of the Role of Inference-Making Skills, Vocabulary Knowledge, Verbal Memory, and Phonological Awareness in the Development of Listening Comprehension From Age 4 to Age 6

264
the children's language, inference-making, and listening comprehension skills were assessed individually from September to November in two sessions, both lasting about 30 minutes. The order of presentation of the tests was the same at time 1 (age 4) and time 2 (age 5). In the first session, we examined children's phonological awareness, vocabulary, and narrative listening comprehension, whereas sentence memory and inference skills were assessed during the second session. At time 3 (age 6), sentence memory, inference skills, and narrative listening comprehension were assessed during the first session, whereas listening comprehension of another narrative text and vocabulary were evaluated during the second session. Children's verbal responses in all tasks, except for rhyming and alliteration, were recorded with an MP3 player, and recordings were used as the basis for scoring.

Phonological Awareness

Children's skill in recognizing words that end or begin with a common sound pattern was examined at age 4 (time 1) by rhyme and alliteration tasks (Silvén, Niemi, & Voeten, 2002). Both tasks included two practice items and 10 test items. The testing was terminated after three successive failures to maintain a child's interest. The testing procedure was adapted from Bryant, MacLean, and Bradley (1990). The Finnish version contained mainly two- and three-syllable nouns because of a lack of usable shorter words in the Finnish language.

The experimenter introduced the rhyming task by singing a familiar nursery song to the child. In the rhyming task, the experimenter showed the child three pictures, simultaneously naming the words they represented. Two of the words rhymed, for instance, pallo (ball), sukka (sock), and kukka (flower). The child's task was to tell the experimenter, by pointing out or vocalizing, the two words that ended with a similar sound pattern. In the alliteration task, the child had to detect which two of the three words named by the experimenter began with the similar sound pattern, for instance, linna (castle), tutti (pacifier), and lintu (bird). Each testing item was scored as correct or incorrect. The total scores for the rhyme and alliteration tasks were the number correct out of 10 items. The coefficients for internal consistency (α) of the rhyming and alliteration tasks were 0.80 and 0.77, respectively. Scores on the rhyme and alliteration variables were correlated (r = .35, p < .001) and were combined into a phonological awareness score (the mean of the scores for the alliteration and rhyme tasks). Silvén and Rubinov (2010) showed that the rhyming and alliteration tasks correlated significantly with syllabic awareness and vocabulary at age 4, which supports the concurrent validity of these tasks.

Vocabulary

The word definition test (Silvén & Rubinov, 2010) was an adaptation of the vocabulary test of the Finnish Wechsler Intelligence Scale for Children (WISC–III). The test was used to assess children's knowledge of word meanings at each of the three timepoints. In the adapted test, the first 10 test words were nouns that were two to four syllables in length, such as kello (watch), satuenvarjo (umbrella), and vasara (hammer). The words became increasingly difficult as the test progressed, and testing was stopped when the child missed three items in a row. There were 32 words total.

The testing procedure was adapted for young children (Silvén & Rubinov, 2010). The words were presented to the child one at a time. The experimenter told the child, “I will say words to you. Listen carefully and tell me what the word means. What is a dog (a practice item)?...What does it look like?...What can it do?...” After the child answered, the experimenter provided the word meaning and the form of the definition: “Yes, the dog is an animal. It is furry. It barks and runs.” If the child did not give any reply for the first test word, the experimenter encouraged the child by repeating the questions. After the practice phase, only a short prompt was used: “Does anything else come to your mind?” Each word definition was scored on a scale of 0–2 in line with the instructions of the WISC–III manual (Wechsler, 1999), but modified according to Snow, Cancino, de Temple, and Schley (1991). The scores were summed to obtain a vocabulary score. A higher score reflected better knowledge of word meanings. The internal consistency reliability (Gronbach’s α) of the standardized test is 0.80 among 6-year-olds (Wechsler, 1999). The internal consistency reliabilities (α) of the adapted test were 0.82 at time 1, 0.81 at time 2, and 0.82 at time 3.
Sentence Memory

The sentence repetition task comprises a subtest of the Developmental Neuropsychological Assessment, which is a standardized test (Korkman, Kirk, & Kemp, 1997). The subtest was used to assess the children’s sentence memory at each of the three timepoints. The child was asked to repeat progressively longer sentences, starting from one word, such as *kukkulauru* (peekaboo), and ending with a 10-word sentence. The experimenter read each sentence aloud and asked the child to recall each sentence immediately. The prompt “Say exactly what I did” was given only for the first two items.

Each trial was scored on a scale of 0–2. According to the manual (Korkman et al., 1997), the child is awarded 2 points for a correctly repeated sentence. One point was given for a repetition containing no more than two errors. Omission or adding of a word, saying a wrong word, and violating the word order were counted as errors. The test included 17 sentences. The assessment was stopped when the child received no points in four consecutive items. The maximum score was 34. The internal consistency (Cronbach’s α) of the test is 0.84 among 4–6-year-old children, and the parallel form correlation is 0.78 (Korkman, 2008). The sentence memory task has been shown to correlate significantly with word span and nonword repetition tasks (Alloway et al., 2004; Korkman, 2000).

Narrative Listening Comprehension

Listening comprehension was assessed by an adaptation of the text comprehension test developed by Silvén and Rubinov (2004) and originally designed for kindergartners and first graders (Vauras, Mäki, Dufva, & Hämäläinen, 1995). We used parallel narrative texts at time 1 (“Misi cat goes hunting”) and time 2 (“Molli & dog is catching”), both describing the behavior of a cat in specific settings (see the Appendix). At time 3, we used the same text that was given at time 1 as well as a parallel text (“*Nalle dog goes stealing*”).

The texts were 91 words long and were made comparable in terms of linguistic properties and macrostructure. The texts consisted of the following sequence of events: character introduction, setting, initiating event, reaction, attempt 1, problem, attempt 2, solution, and outcome reaction. Also, some less relevant information concerning the main topic was included in each narrative.

In the testing situation, the experimenter first introduced the topic of the narrative by saying, “I will read you a story that is about Misi cat. The story tells about when Misi cat was hunting.” Then, the child was instructed to listen carefully to be able to tell about the story afterward. The experimenter read the text aloud twice without stressing any of the main story elements.

Listening comprehension was assessed by a retelling task and four prompted comprehension questions. In the retelling task, the child was asked to tell as much of the story as possible. If the child did not recount anything, the experimenter encouraged the child by a prompt: “You can tell a little about things that happened in the story.” Every child’s retelling was also prompted by asking, “Does anything else come to your mind?”

Children’s retellings and answers to questions were transcribed verbatim. The phrases in the retellings were categorized according to nine narrative elements. Six out of nine elements were based on Mandler and Johnson’s (1977) story grammar framework, and three additional elements were included to get a more detailed picture of the children’s literal story comprehension. One point was given for phrases referring to information from each story element (see the Appendix). Thus, retelling scores ranged from 0 to 9. The percentage agreement method was used to assess inter-rater reliability. Thus, the number of agreements between two independent raters was divided by the total number of responses. Percentage agreements in scoring the retellings were 98% (n = 149) at time 1, 96% (n = 48) at time 2, and 94% (n = 61) at time 3.

To evaluate the sensitivity of this scoring system, we also analyzed retellings at time 1 and time 3 in terms of the propositions the child produced and the accuracy of the child’s comprehension of the semantic idea of the proposition (see Dufva et al., 2001). The scores for story elements at time 1 and time 3 correlated 0.88 and 0.87, respectively, with the scores for proposition analysis, indicating the appropriateness of the scoring based on story elements.

Following the retelling, the story was read again, and then the child was asked four prompted questions to tap his or her understanding of the main events and causal relations (e.g., “About whom did this story tell?” “Why did the cat jump?”). Answers to the questions were scored on a scale of 0–2, yielding a maximum score of 8. For instance, regarding the question “Why did the cat jump?” 2 points were given for an answer referring to the bird and the will to catch (get, hunt) the bird or eat the bird. One point was given for an answer referring to a bird, branch, bush, or jumping. To assess inter-rater reliability, the percentage agreement was calculated for every question. The mean percentage agreements across questions were 98% (n = 149) at time 1, 96% (n = 143) at time 2, and 89% (n = 118) at time 3. To achieve a composite measure of listening comprehension, the retelling and the summed scores for prompted questions (which correlated r = .46 at time 1, r = .53 at time 2, and r = .47 at time 3) were standardized first. Second, the standardized scores were summed and divided by 2. Scores on this task have been found to predict later listening and reading comprehension among preschool and primary school students (Dufva et al., 2001; Lepola, Niemi, Kuikka, & Hannula, 2005).
Inference-Making Skills
We used questions and materials developed by Paris and Paris (2003) to assess children's inference skills through picture book viewing. The picture book Robot-Bot-Bot by Fernando Krahn (1979) was used at each timepoint. The book tells the story of a family (one child and the parents) whose new robot "housecleaner" goes wild after the child plays with its wires. We used an adapted version of the book with 18 pages and one black line drawing on each spread. Prompted implicit questions were used to evaluate inference-making skills (Paris & Paris, 2003).

The experimenter instructed the child by saying,

I will give you a picture book that is about a robot. Look at what is happening in this book. Look carefully at all the pictures, and as you look at the pictures, you may tell me whatever there is going on. Now, you can start.

After the child viewed the picture book, the child was asked to retell as much as possible of the story. The mean times for children viewing the picture book were 159 seconds (SD = 93) at time 1 and 138 seconds (SD = 66) at time 2. After the retelling, the experimenter and the child went through the story together, and the experimenter asked 10 questions. According to Paris and Paris (2003), questions 1–5 are explicit, whereas questions 6–10 are implicit and require inference making to be answered because the information is not directly illustrated in pictures. The data from the last five questions, about the characters' feelings, causal relations, dialogues, predictions, and the theme of the story, were used in the present study. In each question, apart from the thematic question, the child was guided to look at a particular picture (for details, see Paris & Paris, 2003).

Each question was scored on a scale of 0–2. More points were given for an answer when a child integrated information across pages and made connections among the events (Paris & Paris, 2003). To receive 2 points, a child had to refer to at least two other pictures or relate a more global meaning to the picture viewed. For example, for the question asking why the family got the robot, 2 points were given for answers such as "because they don't need to do all the work by themselves" and "because it can then wash and take garbage out." One point was given for a response reflecting page-level information, such as "because no more clean up" or "because it can wash up," and 0 points for no answer or an inappropriate answer, such as "the child wanted to."

Reliability was computed on the basis of the full data set at time 1 and 38% of the sample at time 2. The percentage agreement was calculated for every question. The percentage agreement was above 87% for every question with a mean of 97% for time 1 data and 91% for time 2 data. Inter-rater reliability was also checked by computing Cohen's Kappa for time 1 (0.93) and for time 2 (0.80). A summed score of the five inference questions was calculated.

Results
The results are presented in three sections. First, we report the descriptive statistics for measures of listening comprehension, inference skills, and foundational language variables (i.e., vocabulary knowledge, sentence memory, phonological awareness). In this section, we also report developmental changes on measures of listening comprehension and inference skills as well as concurrent and predictive correlations for the variables. In addition, gender differences in listening comprehension and inference skills are examined to test whether gender should be controlled for in the analyses. This is relevant because by the end of comprehensive school in Finland, differences in literacy achievement between boys and girls are among the largest in the developed world (OECD, 2010, p. 17).

Second, we used structural equation modeling (path analysis with observed variables) (Mplus 5.1; Muthén & Muthén, 2008) to test the fit between our data and a model based on our conceptual model (see Figure 1), as well as determine the direct and indirect paths among the predictors and the outcome variable, kindergarten listening comprehension. Finally, we examined the contribution of concurrent inference and language skills to listening comprehension by means of hierarchical regression analysis.

Descriptive Statistics
Descriptive statistics for the measures at time 1, time 2, and time 3 are shown in Table 1. Possible outliers that might have influenced the results were identified before inspecting the distributions of scores. The scores that were 3 SDs above or below the mean were considered outliers. Only three scores met this criterion: one in vocabulary knowledge at time 2 and two in sentence memory at time 3. The influence of this small number of outliers was considered to be minimal. Table 1 indicates that 14% and 17% of children scored 0 in retelling at time 1 and time 2, respectively. The distribution of the scores for each measure was tested by the Kolmogorov–Smirnov test. The test showed that the data for listening comprehension variables at time 2 and time 3 were normally distributed (p = .20). In addition, this test showed that normality assumption was not met for scores on the foundational language and inference-making variables. However, the distributions of these language and inference skills variables did not substantially differ from normality (skewness below 2 and kurtosis below 5).

To analyze developmental changes and gender differences, we conducted a 3 × 2 (Time × Gender) ANOVA with repeated measures on
listening comprehension and inference skills scores. An ANOVA with linear and quadratic contrasts revealed a significant linear effect of time in retelling elements ($F[1, 128] = 66.81, p < .001, \eta^2_p = 0.34$) and in responses to prompted listening comprehension questions ($F[1, 128] = 165.69, p < 0.001, \eta^2_p = 0.56$). A similar two-way ANOVA for inference skill scores revealed a significant linear effect of time: $F(1, 128) = 215.96, p < 0.001, \eta^2_p = 0.63$. Contrast analyses computed for retelling elements, responses to prompted questions, and inference skills variables further showed significant differences between each successive timepoint. However, no significant gender differences were found either across time in listening comprehension or in inference skills scores. A significant increase was found in both sentence memory ($t[130] = 10.30, p < 0.001$) and productive vocabulary scores ($t[130] = 7.74, p < 0.001$) from time 1 to time 2.

Before reporting results of the path analyses, we describe concurrent and longitudinal correlations of scores on the measures across the three timepoints (see Table 2). The strongest association was found between listening comprehension and sentence memory variables at time 1. Listening comprehension at time 2 was...
more strongly associated with concurrent inference skills and vocabulary knowledge than with sentence memory. A similar pattern of concurrent correlations as found at time 2 was also discovered at time 3. The changing patterns of longitudinal correlations both for listening comprehension (i.e., time 1–2: \( r = .45 \); time 2–3: \( r = .62 \)) and inference skills (i.e., time 1–2: \( r = .41 \); time 2–3: \( r = .55 \)) suggest an increase in the stability of individual differences in those measures across time. In addition, as can be seen in Table 2, the autocorrelations of the variables across time were moderate to strong. Gender was not significantly related to any of the listening comprehension or inference skills variables (\( p > .05 \)), and consequently, it was not included in subsequent analyses.

Consistent with our predictions, inference skills and vocabulary knowledge at time 2 were significantly related to later listening comprehension. Inference skills and vocabulary at time 1 were equally strongly associated with time 2 vocabulary (\( r = .47 \)). Sentence memory at time 1 was only moderately related to later vocabulary knowledge (\( r = .37 \)). Phonological awareness was more strongly correlated with sentence memory at time 2 (\( r = .42 \)) than with listening comprehension at time 2 (\( r = .31 \)). Despite the moderate correlations among the predictor variables, they were not substantially intercorrelated, which decreases the risk of multicollinearity.

### Evaluation of Longitudinal Models via Path Analysis

A structural equation modeling approach to path analysis was used to evaluate the fit of the model to data (Kline, 2011; Schumacker & Lomax, 1996). The conceptual model included 34 parameters. According to Kline, the number of participants to parameter should be 10:1 and should not fall below 5:1. To achieve a better ratio to estimated parameters, the conceptual model was estimated in two phases. Foundational skills were examined in phase 1. The explanatory power of inference skills was tested in phase 2. In other words, there was a model

<table>
<thead>
<tr>
<th>Table 2. Correlations Between Variables Measured at Time 1, Time 2, and Time 3 (N = 130)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>Time 1 (4-year-olds)</td>
</tr>
<tr>
<td>1. Phonological awareness</td>
</tr>
<tr>
<td>2. Vocabulary knowledge</td>
</tr>
<tr>
<td>3. Sentence memory</td>
</tr>
<tr>
<td>4. Inference making</td>
</tr>
<tr>
<td>5. Narrative listening comprehension</td>
</tr>
<tr>
<td>Time 2 (5-year-olds)</td>
</tr>
<tr>
<td>6. Vocabulary knowledge</td>
</tr>
<tr>
<td>7. Sentence memory</td>
</tr>
<tr>
<td>8. Inference making</td>
</tr>
<tr>
<td>9. Narrative listening comprehension</td>
</tr>
<tr>
<td>Time 3 (6-year-olds)</td>
</tr>
<tr>
<td>10. Vocabulary knowledge</td>
</tr>
<tr>
<td>11. Sentence memory</td>
</tr>
<tr>
<td>12. Inference making</td>
</tr>
<tr>
<td>13. Narrative listening comprehension</td>
</tr>
</tbody>
</table>

*p < .01 if |\( r | > .24. **p < .001 if |\( r | > .29.*
without inference skills and a model including inference skills. The specified models were tested using Mplus 5.1 (Muthén & Muthén, 2008). The parameters of the models were estimated using the full information maximum likelihood estimation with nonnormality robust standard errors (MLR estimator; Muthén & Muthén, 2008). MLR estimation was chosen because the distributions of the variables were not normal throughout. Two-tailed statistical testing was used to evaluate the significance of path coefficients (i.e., $t > 1.97$). Based on the standard approach in structural equation modeling, the variables at time 1 were allowed to covary.

The goodness of fit of the estimated models was evaluated by four indicators: the chi-square test, the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR), as described by Hu and Bentler (1999). According to Hu and Bentler, the CFI should be at least 0.95, whereas the RMSEA and the SRMR should be less than 0.08 to indicate an acceptable fit.

**Contribution of Vocabulary, Sentence Memory, and Phonological Awareness to the Development of Narrative Listening Comprehension**

In the first path model, we examined how and to what extent vocabulary, sentence memory, and phonological awareness contributed to the development of listening comprehension. On the basis of theory, the model included autoregressive paths of listening comprehension, vocabulary, and sentence memory. Listening comprehension at time 2 was also assumed to be predicted by vocabulary at time 1 and concurrent sentence memory. Vocabulary at time 2 was assumed to be related to time 3 listening comprehension. We also assumed a path from time 1 sentence memory to time 2 vocabulary. Variation in sentence memory at time 2 was assumed to be accounted for by individual differences in prior sentence memory and phonological awareness.

Although most of the paths were significant (see Figure 2), the fit of model 1 to the data was less than adequate: $\chi^2(df = 13, N = 130) = 56.60, p < .001$, CFI = 0.81, RMSEA = 0.161, SRMR = 0.08. We then sought to identify a better fitting model by testing alternative specifications of the relationships among the variables. Modification indexes suggested that estimation of the path between listening comprehension at time 2 and concurrent vocabulary, as well as the path between listening comprehension at time 1 and listening comprehension at time 3, would increase the fit of the model significantly. The decision to model the concurrent path from vocabulary to listening comprehension was supported by cross-sectional studies (Florit et al., 2009; Kendeou et al., 2008). However, our decision to model the link between listening comprehension at time 1 and

**Figure 2. Path Analysis for Model 1: Foundational Language Skills as Predictors of Narrative Listening Comprehension**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Time 1 (4-year-olds)</th>
<th>Time 2 (5-year-olds)</th>
<th>Time 3 (6-year-olds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrative listening comprehension</td>
<td></td>
<td>.32</td>
<td>.54</td>
</tr>
<tr>
<td>Vocabulary knowledge</td>
<td>.15</td>
<td></td>
<td>.16</td>
</tr>
<tr>
<td>Sentence memory</td>
<td>.39</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td>Phonological awareness</td>
<td>.22</td>
<td>.23</td>
<td></td>
</tr>
</tbody>
</table>

Note. Standardized path coefficients were estimated by Mplus (amount of variance explained is displayed in parentheses). All paths displayed as solid lines are significant at $p < .05$ (two-tailed).
listening comprehension at time 3 was admittedly ad hoc and suggested by empirical data rather than theory.

First, model 2 was specified to test the concurrent effect of vocabulary to listening comprehension at time 2. The fit indexes for model 2 were also not acceptable: \( \chi^2(\text{df} = 12, N = 130) = 31.57, p = .001, \text{CFI} = 0.92, \text{RMSEA} = 0.112, \text{SRMR} = 0.058. \)

Next, model 3 was specified by adding a direct path from listening comprehension at time 1 to listening comprehension at time 3. This path suggests that time 1 includes a component that is not developing through time 2 but nevertheless is important in the prediction of listening comprehension at time 3. The fit of model 3 to the data was acceptable: \( \chi^2(\text{df} = 11, N = 130) = 17.59, p = .092, \text{CFI} = 0.97, \text{RMSEA} = 0.068, \text{SRMR} = 0.042. \)

All the paths that were nonsignificant in model 3 were removed. Thus, the path from vocabulary at time 1 to listening comprehension at time 2 (\( t = 0.12 \)), the path from vocabulary at time 2 to listening comprehension at time 3 (\( t = 1.37 \)), and the path from sentence memory at time 2 to listening comprehension at time 2 (\( t = 0.49 \)) were omitted. The chi-square difference test showed that deleting these paths from model 3 did not significantly decrease the fit of the model to the data: \( \Delta \chi^2(3) = 2.09, p = .554 \). In addition, phonological awareness at time 1 and sentence memory at time 2 were omitted from the subsequent model because neither contributed directly or indirectly to listening comprehension.

After the above modifications, the resulting model, model 4, showed a good fit: \( \chi^2(\text{df} = 6, N = 130) = 6.55, p = .364, \text{RMSEA} = 0.027, \text{CFI} = 0.99, \text{SRMR} = 0.028. \) This model with standardized path coefficients and squared multiple correlations is shown in Figure 3. All paths depicted in Figure 3 were statistically significant.

As can be seen from Figure 3, the proportion of total variance in kindergarten listening comprehension accounted for by these paths was 44%. The results showed that vocabulary knowledge did not contribute directly to later listening comprehension at time 3 when the direct effects of listening comprehension at time 1 and time 2 were taken into account. However, children’s vocabulary at age 5 had a unique concurrent effect on listening comprehension. Vocabulary knowledge at time 2 was predicted both by individual differences in previous vocabulary and sentence memory. These predictors accounted for 26% of the total variance in vocabulary knowledge.

To examine mediated paths from time 1 and time 2 measures to children’s listening comprehension at time 3, the total effects for each measure were parsed into direct and indirect influences. Indirect effect can be measured by the product of the standardized coefficients that constitute each indirect path (Schumacker & Lomax, 1996). Model 4 provided support for the indirect contribution of vocabulary and sentence memory to later listening comprehension. The developmental relationship between vocabulary at time 1 and listening comprehension at time 3 was mediated through vocabulary at time 2 and through listening comprehension at time 2 (standardized indirect effect = 0.08, \( t = 3.33, p = .001 \)). Sentence memory at time 1 was indirectly

---

**Figure 3. Path Analysis for Model 4: Foundational Language Skills as Predictors of Narrative Listening Comprehension**

<table>
<thead>
<tr>
<th>Time 1 (4-year-olds)</th>
<th>Time 2 (5-year-olds)</th>
<th>Time 3 (6-year-olds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrative listening comprehension</td>
<td>Narrative listening comprehension (35%)</td>
<td>Narrative listening comprehension (44%)</td>
</tr>
<tr>
<td>Vocabulary knowledge</td>
<td>Vocabulary knowledge (26%)</td>
<td></td>
</tr>
<tr>
<td>Sentence memory</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Standardized path coefficients were estimated by Mplus (amount of variance explained is displayed in parentheses). All paths displayed as solid lines are significant at \( p < .05 \) (two-tailed).
related to listening comprehension at time 3 through vocabulary at time 2 and through listening comprehension at time 2 (standardized indirect effect = 0.05, t = 2.00, p = .045). These analyses indicated that vocabulary and sentence memory are important resources involved in narrative listening comprehension, yet they do not directly contribute to later listening comprehension over and above the autoregressor.

**The Relations of Inference and Foundational Language Skills to the Development of Narrative Listening Comprehension**

In the second stage of modeling, we analyzed the developmental relations among inference skills, listening comprehension, vocabulary, and sentence memory in children from the age of 4 to 5 years and examined the contribution of these variables to listening comprehension at age 6.

We started to examine model 5 by including the same concurrent and longitudinal paths that were found to be significant in model 4. Model 5 contained one additional autoregressive path, from time 1 inference skills to time 2 inference skills, and four cross-lagged relationships: (1) from time 1 inference skills to time 2 listening comprehension, (2) from time 1 listening comprehension to time 2 inference skills, (3) from time 2 inference skills to time 3 listening comprehension, and (4) from time 1 inference skills to time 2 vocabulary. Variability in listening comprehension at time 3 was assumed to be directly predicted by inference skills and listening comprehension at time 2, as well as by listening comprehension at time 1.

The statistics indicated that the fit of this model to the data was not adequate: \( \chi^2(\text{df} = 11, N = 130) = 25.04, p = .009, \text{RMSEA} = 0.099, \text{CFI} = 0.934, \text{SRMR} = 0.06. \)

However, modification indexes suggested that estimation of the concurrent path from inference skills at time 2 to listening comprehension would increase the fit of the model. The direction of this path was based on studies suggesting a causal contribution of inference skills to comprehension (Oakhill & Cain, 2011; Paris & Paris, 2007). After adding this path to the model and removing the nonsignificant path with the lowest t-value (i.e., the link between inference skills at time 1 and listening comprehension at time 2), model 6 fit the data well: \( \chi^2(\text{df} = 11, N = 130) = 15.70, p = .153, \text{RMSEA} = 0.057, \text{CFI} = 0.978, \text{SRMR} = 0.045. \) Model 6 with standardized path coefficients and squared multiple correlations is shown in Figure 4.

As Figure 4 shows, inference skills at time 2 contributed uniquely to subsequent listening comprehension even after taking into account the autoregressors at time 1 and time 2. When the effects of inference skills were modeled, vocabulary knowledge was the only language skill contributing indirectly to

---

**Figure 4. Path Analysis for Model 6: Foundational Language Skills as Predictors of Narrative Listening Comprehension**

![Path Analysis Diagram](image)

**Note.** Standardized path coefficients estimated by Mplus (amount of variance explained is displayed in parentheses; covariances among time 1 variables are not shown). All paths displayed as solid lines are significant at \( p < .05 \) (two-tailed).
kindergarten listening comprehension (see Table 3). The results also showed that both time 1 inference skills and time 1 listening comprehension accounted for a significant amount of variance in children’s ability to make inferences at time 2 (24%). Vocabulary knowledge at time 2 was predicted by previous vocabulary, as well as by inference skills measured at time 1. In fact, inference skills, vocabulary, and sentence memory measured at time 1 together accounted for 32% of variance in vocabulary at time 2.

Both vocabulary and inference skills at time 2 had concurrent effects on listening comprehension, above and beyond the autoregressor. Individual differences in listening comprehension at time 1 were directly, as well as indirectly through listening comprehension at time 2, associated with individual differences in listening comprehension at time 3. The total amount of variance in kindergarten listening comprehension accounted for by previous listening comprehension, inference skills, and vocabulary knowledge was 45%. In addition, several indirect paths from time 1 measures and from time 2 measures to the outcome measure were also significant (see Table 3).

Inference skills at time 1 were indirectly associated with listening comprehension at time 3, predominantly as a result of indirect paths either through time 2 inference skills or through time 2 vocabulary scores and through time 2 listening comprehension (see Table 3). However, the indirect effect of time 1 inference skills through time 2 inferences on listening comprehension only approached statistical significance ($p = .065$). The total effect of time 2 inference skills on time 3 listening comprehension was accounted for by the direct relationship ($\beta = 0.17, p = .028$) and by the indirect path through time 2 listening comprehension ($\beta = 0.11, p = .005$). As Table 3 shows, the significant relationship between time 1 listening comprehension and time 3 listening comprehension was also mediated by inference skills at time 2. In fact, the indirect links through inference skills accounted for 19% (i.e., 0.08/0.43) of the total effect of initial listening comprehension on the outcome measure. These findings suggest that listening comprehension and inference skills are contributing to one another across time in the prediction of subsequent listening comprehension. As assumed, time 1 vocabulary knowledge was significantly indirectly related to listening comprehension through time 2 vocabulary.

## Concurrent Contributions of Inference Skills to Narrative Listening Comprehension

Finally, a hierarchical regression analysis was conducted to examine the unique contribution of concurrent

### Table 3. Standardized Total, Direct, and Indirect Effects on Listening Comprehension for the Final Model

<table>
<thead>
<tr>
<th>Path</th>
<th>Effect estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>From IS at time 1 to LC at time 3</td>
<td>.13**</td>
</tr>
<tr>
<td><strong>Total effects</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Indirect effects</strong></td>
<td></td>
</tr>
<tr>
<td>IS at time 1 →</td>
<td></td>
</tr>
<tr>
<td>IS at time 2 → LC at time 3</td>
<td>.05†</td>
</tr>
<tr>
<td>IS at time 2 → LC at time 2 → LC at time 3</td>
<td>.04*</td>
</tr>
<tr>
<td>VK at time 2 → LC at time 2 → LC at time 3</td>
<td>.04*</td>
</tr>
<tr>
<td>Total effects</td>
<td>.28**</td>
</tr>
<tr>
<td><strong>Direct effects</strong></td>
<td></td>
</tr>
<tr>
<td>IS at time 2 →</td>
<td></td>
</tr>
<tr>
<td>LC at time 2 → LC at time 3</td>
<td>.11**</td>
</tr>
<tr>
<td>From LC at time 1 to LC at time 3</td>
<td></td>
</tr>
<tr>
<td><strong>Total effects</strong></td>
<td>.43**</td>
</tr>
<tr>
<td><strong>Indirect effects</strong></td>
<td></td>
</tr>
<tr>
<td>LC at time 1 →</td>
<td></td>
</tr>
<tr>
<td>LC at time 2 → LC at time 3</td>
<td>.09**</td>
</tr>
<tr>
<td>IS at time 2 → LC at time 3</td>
<td>.05†</td>
</tr>
<tr>
<td>IS at time 2 → LC at time 2 → LC at time 3</td>
<td>.03*</td>
</tr>
<tr>
<td>Total effects</td>
<td>.04**</td>
</tr>
<tr>
<td><strong>Indirect effects</strong></td>
<td></td>
</tr>
<tr>
<td>VK at time 1 →</td>
<td></td>
</tr>
<tr>
<td>VK at time 2 → LC at time 2 → LC at time 3</td>
<td>.04**</td>
</tr>
<tr>
<td>From SM at time 1 to LC at time 3</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>.02</td>
</tr>
<tr>
<td><strong>Indirect effects</strong></td>
<td></td>
</tr>
</tbody>
</table>
| SM → VK at time 2 → LC at time 2 → LC at time 3 | .02

Note. IS = inference skills. LC = narrative listening comprehension. SM = sentence memory. VK = vocabulary knowledge. $† p < .10. * p < .05. ** p < .01.$
The results showed that inference-making skills contributed both directly and indirectly, through vocabulary knowledge, to later narrative listening comprehension. This is in line with our hypotheses and with previous research (Florit et al., 2011; Kendeou et al., 2008; Oakhill & Cain, 2011). Also as predicted, the contribution of narrative listening comprehension at age 4 to listening comprehension at age 6 was mediated by inference-making skills at age 5.

On the contrary, the results regarding the role of foundational skills were not entirely consistent with our predictions. Vocabulary did not predict later narrative listening comprehension beyond the autoregressive effect of listening comprehension. However, vocabulary was significantly associated with concurrent narrative listening comprehension at age 5. Also contrary to our hypothesis, the data did not support a developmental link from sentence memory at age 4 to listening comprehension at age 6. More specifically, this link was nonsignificant after the inclusion of inference skills in the model. Sensitivity to the phonological structure of words at age 4 was also not related to the development of narrative listening comprehension, contrary to our hypothesis.

Overall, the results based on our final path model showed that listening comprehension, inference skills, and vocabulary measured at 4 and 5 years of age and sentence memory measured at age 4 accounted for 45% of variance in listening comprehension at age 6. In comparison with the path model with only foundational language skills, the total amount of variance in listening comprehension explained at age 6 was only marginally larger in the path model with inference skills. However, inference making was a potent text-level skill. A sizable proportion of the variance (49%) in listening comprehension at age 6 was explained by the autoregressor, as well as concurrent vocabulary, sentence memory, and inference skills. It is worth noting that inference skills explained 6% of additional variance over and above the autoregressor and concurrent language resources.

**Development of Inference Skills and Narrative Listening Comprehension**

The present findings add to previous knowledge of the role of inference making in the development of comprehension (Gain & Oakhill, 1999; Gain, Oakhill, & Bryant, 2004; Pike, Barnes, & Barron, 2010) by confirming that inference skills are already in place at the age of 4 years and are important in narrative listening comprehension. Inference skills made a unique contribution to the development of listening comprehension and vocabulary, a new finding which fits with multicomponent models of text comprehension (Kintsch & Rawson, 2005; Oakhill & Cain, 2007; Perfetti et al., 2005). This finding is unique among the literature on listening comprehension.
comprehension in prereaders because we also included the autoregressor (i.e., listening comprehension at an earlier timepoint) in the model. Specifically, path modeling suggested that the more skilled children were at age 5 in making inferences after picture book viewing, the better they could retell a story and answer questions about it at age 6. The finding is suggestive of a causal relation between inference skills and listening comprehension, which supports the results of other studies, particularly those showing the benefits of inference-making training for listening comprehension (Paris & Paris, 2007) and reading comprehension (Yuill & Oakhill, 1988).

An interesting result was that narrative listening comprehension at age 4 was a significant predictor of inference skills at age 5, which, in turn, directly predicted narrative listening comprehension at age 6. These findings suggest a reciprocal relationship between listening comprehension and inference skills through development. These findings are analogous to those of Oakhill and Cain (2011), who showed a similar relationship between reading comprehension and inference skills from age 7 to age 11.

Our path analysis and hierarchical regressions also revealed that the predictive strength of inference skills increased with age. A tentative explanation is that picture-level inferences dominate at age 4, whereas later inferences tend to span across pictures (i.e., across different events in the narrative) and therefore have a stronger relationship to overall comprehension. Like other studies that have explored comprehension processes across different modes of presentation (e.g., Kendeou et al., 2008), we found evidence that children use similar text-level skills when comprehending pictorial and oral narratives. Consequently, our results support the view that inferential processes involved in the acquisition of a meaning-based representation of a story are not specific to oral and written language but are at work in the comprehension of narratives presented in other media (Gernsbacher, 1990).

The finding of the direct effect of listening comprehension at age 4 on listening comprehension at age 6 was at odds with our assumption that the autoregressor at age 5 would be sufficient to predict subsequent listening comprehension at age 6. Consequently, we tested whether the direct link found between time 1 and time 3 listening comprehension could have been caused by the fact that the time 1 text also served as one of the texts used at time 3. We tested this hypothesis by omitting the time 1 text at time 3. However, both time 1 and time 2 listening comprehension contributed significantly to the prediction of time 3 listening comprehension. Thus, the direct link between time 1 and time 3 listening comprehension may reflect temporary changes in the growth of comprehension. These differences may also be determined by external influences that are not captured at time 2. Such influences could be, for instance, less shared book reading at later ages compared with age 4 or variability in children's interest in narratives.

**Development of Foundational Skills and Their Role in Listening Comprehension**

The results also clarified the developmental dynamics among different foundational language skills, inference making, and narrative listening comprehension. In particular, our findings shed light on the development of the relationship between inference making and vocabulary knowledge. The finding that inference making at age 4 was a unique predictor of vocabulary knowledge at age 5 over and above the effects of vocabulary and sentence memory at age 4 is important. Analogous results have been reported by Cain, Oakhill, and Elbro (2003), who showed that children's ability to derive the meaning of words from context correlates with their reading comprehension. In line with this, Gathercole et al. (1992) suggested that relative to phonological memory, a child's ability to derive meaning from context might gradually become more important in vocabulary learning. In fact, Nash and Snowling (2006) found in children with poor vocabulary that a strategy to infer the meaning of new words from the written context was more effective in increasing expressive vocabulary and reading comprehension than teaching definitions of the words.

It is worth noting that unlike most studies of listening comprehension, expressive vocabulary was used in the present study. Tasks based on expressive vocabulary are more challenging than those based on receptive vocabulary, especially for young children. According to Snow et al. (1991), when a child is asked to tell what a word means, the child has to analyze the meaning of the target word against other candidates and then organize the focal meaning into a definition. Thus, inference-making questions may not only draw on similar higher level conceptual skills as those demanding expressive vocabulary. In addition, a child’s inference making in response to a narrative may also enrich his or her semantic representation of the relatedness of words.

Because spoken words function as a principal carrier for meaning, knowledge of word meanings has at least a facilitating role in listening comprehension (Pearson, Hiebert, & Kamil, 2007). In our path models, we showed that better knowledge of word meanings did not directly contribute to later comprehension but was significantly associated with the concurrent listening comprehension. This is partly in contrast to findings by Sénéchal et al. (2006) but in line with the results of studies by Florit et al. (2011) and Kendeou et al. (2008). We found that vocabulary and sentence memory did not serve as unique predictors of
later listening comprehension probably because of the shared variance that these language skills had with more potent predictors such as prior listening comprehension and inference skills. Thus, the difference in the findings of the present study and those of Sénéchal et al. may be due to the fact that we not only controlled for the autoregressive effect of listening comprehension but also included inference skills in the longitudinal model. However, it is worth noting that vocabulary was consistently and significantly associated with listening comprehension scores across the three timepoints, suggesting that the narrative texts we used required children to draw considerably on their vocabulary knowledge.

We also found that after the contribution of the autoregressors and vocabulary, time 3 sentence memory accounted for a small (2%) but significant amount of variance in listening comprehension at time 3 (see also Florit et al., 2009). Sentence memory was also strongly related to concurrent listening comprehension at age 4. Thus, although sentence memory did not play a predictive role in this study, we should not overlook the role of verbal working memory in the development of a child’s listening comprehension.

Phonological awareness was associated neither with the development of listening comprehension nor the development of inference skills. Although this was not in line with our prediction that phonological awareness would play an indirect role in narrative listening comprehension via sentence memory, the finding is in line with previous research that suggested that inference skills and listening comprehension develop relatively independently from phonological skills (Dufva et al., 2001; Kendeou et al., 2008; Lynch et al., 2008). In fact, previous studies have been quite unanimous in their conclusion that children’s phonological awareness (i.e., emerging ability to focus on word forms instead of word meanings) is a major predictor of word reading and writing acquisition (e.g., Mäki, Voeten, Vauras, & Poskiparta, 2001; Silvén et al., 2007).

Oral language skills, including vocabulary, have been found to predict phonological skills rather than vice versa (NICHD, 2005; Storch & Whitehurst, 2002). Although both oral language and phonological awareness have been found to lay the foundation for word reading acquisition (Kendeou et al., 2009), oral language comprehension seems to be a more potent predictor of later reading comprehension (de Jong & van der Leij, 2002; Lepola et al., 2005).

**Limitations and Directions for Future Research**

The present study has some apparent limitations. We did not assess verbal IQ, which has been controlled for in other longitudinal studies of text comprehension (Dufva et al., 2001; Oakhill & Gain, 2011). We used an expressive vocabulary test, based on the WISC–III, as an approximation of verbal IQ instead of the full scale. We also did not control for measurement error in the assessment of narrative listening comprehension. In addition, only four prompted questions were presented, and they assessed information explicitly given in the stories. However, the scores obtained from the two scoring systems applied to the retellings (i.e., analyses for narrative elements and propositions) were strongly correlated, which provides some support for the reliability of our listening comprehension measure. We also acknowledge that the growth observed in inference skills may be confounded by learning effects caused by exposure to the same picture book across time. Our decision to use the same task was based on the study by Paris and Paris (2003) that showed no significant practice effects. Also, we note that we modified our conceptual model by deleting several paths and adding three. This was done on empirical rather than theoretical grounds. To generalize our findings, the final model should be tested with other data sets.

It is possible that other factors not included in this study contributed to the development of children’s narrative listening comprehension, such as comprehension monitoring (Kinnunen, Vauras, & Niemi, 1998; Skarakis-Doyle, 2002), morphological skills (Silvén et al., 2007), children’s exposure to storybooks (Sénéchal & LeFevre, 2002), and the quality of scaffolding during joint picture book reading (Silvén, Ahtola, & Niemi, 2003). We also acknowledge the potential role of other resources that children might use to construct a mental representation of a narrative text, such as world knowledge (Best, Floyd, & McNamara, 2008) and knowledge of story structure (Stevens, Van Meter, & War cholak, 2010). These factors should be considered in future studies.

The scope of the current study is limited to the development of listening comprehension in the preschool and kindergarten years. Thus, we cannot make direct conclusions about how the developmental paths found in this study relate to reading acquisition and later reading achievement. However, based on theoretical conceptualizations as well as empirical evidence (e.g., Hoover & Gough, 1990; Kendeou et al., 2009; McKeown & Beck, 2006; Proctor, Carlo, August, & Snow, 2005; Verhoeven & Leeuwe, 2008), listening comprehension skills are an important precursor to the development of reading comprehension. As such, our findings add to the literature on preliteracy skills, providing more detailed information than previously available on inference making and its role in listening comprehension. Future studies should explore more directly how the development of inference-making, listening comprehension, and foundational skills in the preschool years...
The Role of Inference Making and Other Language Skills in the Development of Narrative Listening Comprehension in 4–6-Year-Old Children

Contributions to Reading Achievement in the Elementary School Years.

Implications for Assessment and Theoretical Models for Comprehension

The findings of the present study have practical implications in terms of the assessment of comprehension skills, and theoretical implications in terms of models of the development of narrative listening comprehension. The results suggest that the assessment of listening comprehension among prereaders should focus not only on children's retelling abilities and basic language skills, such as vocabulary, but also on children's inference-making skills. The present data suggest that the assessment of inference skills by implicit questions after reading a picture book provides a viable tool for teachers as well as speech and language therapists to monitor the progress of the child in meaning making. This type of assessment may also help professionals evaluate the child's strengths as well as identify possible difficulties in story comprehension that extend beyond verbal memory and vocabulary.

The variability of scores and the absence of floor and ceiling effects in the inference-making measures in this study lend support for the developmental validity and diagnostic sensitivity of wordless picture book tasks for children who are not yet reading. These tasks are not only applicable for preschool- and kindergarten-age children but also for older children. These tasks would allow educators to monitor the growth of comprehension using an assessment that is not confounded by decoding skill. It is important to consider the difficulty of the text in the assessment of comprehension abilities. An early study by Allen (1985) showed that the level of inferential comprehension was strongly related to the level of decoding when the students read a narrative written by an adult. However, when a narrative was based on the child's own language and experiences, decoding was not related to inferential comprehension.

The results support theoretical models of narrative comprehension (Kintsch, 1998; Perfetti, 1999). Foundational language skills, namely vocabulary and verbal memory, are necessary for narrative comprehension but insufficient by themselves. In our results, these foundational language skills measured concurrently tended to relate to narrative listening comprehension, but they did not directly predict later listening comprehension after controlling for the autoregressor. Text-level cognitive processes, including inference making, are also necessary, but insufficient, for the development of listening comprehension. Inference-making skills allow a comprehender to integrate the information in text or pictures to form a coherent mental representation, or situation model, of the narrative (Kintsch, 1998).

As would be predicted based on theoretical conceptualizations, we found longitudinal relations between inference making and listening comprehension. Furthermore, like previous studies, we found evidence of a reciprocal relationship between listening comprehension and inference making. Not only does inference making facilitate listening comprehension, but as young children gain more experience with listening comprehension, they also become more familiar with the causal structure of narratives, likely facilitating their inference-making skills (Lynch et al., 2008).

Our findings also clarify the relationships between foundational skills and inference making. Vocabulary and verbal memory are necessary for inference making within narratives. Children need to understand words and have mental resources to connect different ideas to make inferences within narratives. Simultaneously, inference making may well facilitate vocabulary acquisition, as children learn to infer the meanings of words from context. Our findings support the theorized reciprocal relationship between vocabulary and inference-making skills, although the facilitating role of verbal memory is less clear. Overall, the contribution of inference-making skills to the development of narrative listening comprehension from age 4 to age 6 indicates that inference skills are already in place among prereaders and play an important role in the construction of a meaning-based representation of a narrative.

Notes

The study is based on a longitudinal research project (Promo) conducted at the University of Turku, Finland, and supported by a grant (No. 118330) from the Academy of Finland to the first author. We wish to thank the participating children as well as the administrators and teachers of the day-care centers where the follow-up study took place. The first author also wishes to thank the University of Michigan and the pathways group led by Professor Frederick J. Morrison in providing a supportive milieu to write the manuscript. We are also grateful to Dr. Marja-Kristiina Lerkkanen for her valuable comments.

References


Literature Cited


Submitted June 10, 2011
Final revision received February 1, 2012
Accepted February 17, 2012

Janne Lepola is a researcher at the University of Turku, Rauma, Finland; e-mail janlep@utu.fi.

Julie Lynch is an associate professor at Saginaw Valley State University, University Center, Michigan, USA; e-mail jslynch@svsu.edu.

Eero Laakkonen is a statistician at the University of Turku, Finland; e-mail eerlaa@utu.fi.

Maarit Silvén is a professor at the University of Turku, Rauma, Finland; e-mail maarsi@utu.fi.

Pekka Niemi is a professor at the University of Turku, Finland; e-mail peknie@utu.fi.
Appendix

Misi Cat Narrative Text and Examples of Scoring for the Elements From Children’s Retellings

Narrative Text

The narrative text in the Table (Vauras & Friedrich, 1994) was categorized according to nine elements based on story grammar categories as well as event sequences referring to reactions, the goal and attempts of the character. One point was given for phrases indicating the presence of each story element.

Table. The Narrative Misi Cat Text and the Elements Used to Score Children’s Retellings

<table>
<thead>
<tr>
<th>The elements</th>
<th>The narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character and setting</td>
<td>At a summer cottage, Misi cat likes best to go out hunting. She often goes strolling along paths and ditches.</td>
</tr>
<tr>
<td>Initiating event</td>
<td>One day Misi is walking in the yard and hears rustling in the nearby grass.</td>
</tr>
<tr>
<td>Reaction 1</td>
<td>Misi stops, alerted, and pricks up her ears.</td>
</tr>
<tr>
<td>Goal/attempt 1</td>
<td>She starts to creep cautiously toward the prey while pressing herself tightly against the ground.</td>
</tr>
<tr>
<td>Problem</td>
<td>But a bird in the grass suddenly flies up on a branch of a high shrub.</td>
</tr>
<tr>
<td>Attempt 2</td>
<td>Misi cat presses her hind legs strongly on the ground and jumps a wild leap toward the branch.</td>
</tr>
<tr>
<td>Solution</td>
<td>But at the last moment, the bird soars into the air and escapes.</td>
</tr>
<tr>
<td>Reaction 2</td>
<td>Misi starts washing herself vehemently to forget the embarrassing incident. She licks her front paws long and carefully. Misi is, however, very fond of hunting during summertime.</td>
</tr>
</tbody>
</table>

Scoring Criteria and Examples

Character

1 point: The cat, the bird, or both are mentioned. Nicknames for the cat (e.g., kisi [kitty], misu [pussy]) or species of animal (e.g., crow) are accepted. (Examples: “It happened in the fairy tale that the cat was hunting, and the cat and the bird flew.” “That Misi cat hunted.” “Pussy is hunting.” “The crow jumped and took flight.”)

Setting

1 point: Either where the cat is or what she is doing is referred to as an answer. (Examples: “The cat was prey-ing.” “She was hunting.” “First, she was in the summer cottage.” “She was let out, and then the cat left the cottage.” “The cat was walking in the woods and along the ditches.”)

Initiating Event

1 point: The cat hearing a rustling/sound/chirp/noise is referred to as an answer. (Examples: “There was some noise coming from the grass, and that was the bird.” “She ran quickly to see where the noise was coming from, and it happened to be the bird.” “She saw something in the grass.”)

Reaction 1

1 point: It is evident from the answer how Misi reacts to the noise in the grass; she stops, alerted, and pricks up her ears. (Examples: “She pricks up her ears.” “The pussy came to see and pricked her ears.”)

Goal/Attempt 1

1 point: The answer shows that the cat tried to catch the prey by pressing herself against the ground/ by creeping cautiously. (Examples: “And then she pressed herself against the ground.” “Crept cautiously toward the prey.”)

Problem

1 point: It is mentioned in the answer that the bird flies to the branch/high/to the tree/to the shrub. (Examples: “She flew to the branch of the tree.” “As the bird went into the shrubbery, to the high branch of the shrub.” “And luckily the bird flew on the branch.”)

0 points: No scores are given for an answer that clearly indicates that the episode has been misunderstood.
or an answer that is said unclearly. (Examples: “That the cat went and climbed to the shrubbery to hide.” “There was a nest on the branch, and there was no others.”)

**Attempt 2**

1 point: The child has understood that the cat made an effort/jumped to get the bird from the branch. (Examples: “And then she jumped.” “Misi jumps on the branch.” “The kitty tried hard and hard and put her legs on the ground and so jumped on the branch.” “She tried; she set her paws steady on the ground. And then she ran up and jumped.”)

**Solution**

1 point: It is evident from the answer that the bird escapes. (Examples: “The bird escaped.” “The bird flew away.” “The crow jumped and escaped.” “Then she couldn’t get it because she flew away.”)

**Reaction 2**

1 point: The answer describes the cat’s reaction to the fact that the bird escaped: she licked herself (to forget the embarrassing incident). (Examples: “Then she… licked her paw.” “She is going to wash herself.” “Then she licked, the cat did.”)