



OPEN ACCESS

EDITED BY

Gordana Hrzica,
University of Zagreb, Croatia

REVIEWED BY

Kristine Marie Jensen De López,
Aalborg University, Denmark
Sveta Fichman,
Hadassah Academic College, Israel

*CORRESPONDENCE

Josefin Lindgren
✉ josefin.lindgren@nordiska.uu.se

RECEIVED 03 July 2023

ACCEPTED 18 September 2023

PUBLISHED 11 October 2023

CITATION

Lindgren J (2023) Age and task type effects on comprehension and production of narrative macrostructure: storytelling and retelling by Swedish-speaking children aged 6 and 8. *Front. Commun.* 8:1252260. doi: 10.3389/fcomm.2023.1252260

COPYRIGHT

© 2023 Lindgren. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Age and task type effects on comprehension and production of narrative macrostructure: storytelling and retelling by Swedish-speaking children aged 6 and 8

Josefin Lindgren*

Department of Scandinavian Languages, Uppsala University, Uppsala, Sweden

It is often said that story retelling tasks, where children listen to a model story and then retell it, are easier than story telling tasks, where children are not provided with a model. However, previous studies have rarely used comparable stimuli and procedures for the different tasks, creating possible confounds with task effects. Additionally, studies seldom investigate the interaction between age and task type and most studies focus on preschool children. The present study addresses these gaps by analyzing the performance of Swedish-speaking 6-year-olds and 8-year-olds ($N = 74$) on measures of story comprehension and story structure (narrative macrostructure) using a carefully controlled procedure with comparable telling and retelling tasks (MAIN Cat/Dog stories) and counterbalancing the order of the tasks. For story comprehension, results showed that overall accuracy was uniformly high (>90%) across tasks and age groups. However, performance was substantially lower for one question (D10), which assesses comprehension of the entire plotline. With increasing age, children did not become more likely to answer this question correctly, nor did hearing a model story improve performance. A qualitative analysis showed that incorrect answers often contained reasonable explanations showing advanced general inferencing abilities. In light of these results, an adjustment to the scoring of MAIN is recommended. For story structure, results showed significant effects of both age and task type, with higher scores in retelling and higher scores by the 8-year-olds. The 8-year-olds exhibited the same performance gap between telling and retelling as the 6-year-olds. There was also a significant effect of task order, showing a training effect from the first task to the second. The present study thus confirms previous findings that expressing story structure in a retelling task is easier than in a telling task, but showing this for the first time while controlling for task order and stimulus complexity in MAIN.

KEYWORDS

macrostructure, Multilingual Assessment Instrument for Narratives (MAIN), narratives, story comprehension, story retelling, story structure, storytelling, Swedish

1. Introduction

Picture-based narrative tasks offer valuable insights into children's language development. They make it possible to investigate various aspects of language (see e.g., Pavlenko, 2008) in language samples that are relatively comparable across individuals and simultaneously offer an assessment of children's language skills that is ecologically valid (e.g., Botting, 2002). These features, as well as the increased availability of standardized narrative assessment instruments, such as the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015, 2019b), for a large number of languages, have made elicited narratives an increasingly popular method in language acquisition research in recent years. Different procedures or task types can be used to elicit narratives. Two main types of narrative tasks are *retelling* and *telling* (story generation) tasks. In a retelling task, participants first listen to a story, either with or without visual support (such as pictures or a video clip), after which they are asked to retell it to a listener. In a telling (story generation) task, participants are asked to tell a story without having been provided with an oral model. Instead, they are either given a simple oral prompt (such as "tell me a story about a frog"), or are asked to tell the story shown in a single picture, a sequence of pictures or in a video clip. It is often said that story retelling tasks are easier for children than story telling tasks (e.g., Liles, 1993; Lever and Sénéchal, 2011), since in retelling tasks, children can draw on content provided in the oral model. Indeed, previous studies indicate that children show better performance on measures of the overarching structure of the narrative, the *narrative macrostructure* or *story structure*, in retelling tasks than in telling tasks (see Section 2). However, these studies have often used telling and retelling tasks that are not comparable to each other due to differences in the stimuli and/or have not controlled for the order in which the two tasks have been administered. The majority of studies comparing retelling and telling have also investigated children aged 4 to 6; studies offering comparisons with older children, as is done in the present study, can provide additional evidence as to how narrative competence in retelling and telling tasks develops during the early school years. Additionally, the interaction between age and task type, that is, whether children of different ages perform differently on telling and retelling tasks, has rarely been investigated.

In addition to analyzing narrative production in terms of story structure, assessing *story comprehension* provides valuable information about the child's ability to understand the relationships between events, such as reasons for characters' emotions and actions. Story comprehension is often measured based on answers to comprehension questions targeting different aspects of the narrative macrostructure, such as characters' goals. Studies comparing performance on telling and retelling tasks for both story comprehension and story structure are relatively rare. The present study seeks to address these gaps by comparing the performance of Swedish-speaking children in two age groups (6-year-olds, 8-year-olds) on comparable telling and retelling tasks, analyzing measures of both comprehension and production of narrative macrostructure (story structure). To enable comparison with previous research, the present study uses the Multilingual Assessment Instrument for Narratives

(hereafter MAIN; Gagarina et al., 2012, 2015, 2019b), a picture-based narrative instrument developed for children aged 4 to 10 with standardized elicitation procedures for both a story telling and a story retelling task, as well as scoring protocols for story comprehension and story structure (narrative macrostructure).¹ MAIN is currently available in over 80 languages and has been used in a large number of studies covering a wide range of languages and language combinations (for a recent overview of studies using MAIN to investigate story comprehension and story structure, see Lindgren et al., 2023).

The paper proceeds as follows. First, results from studies investigating effects of age and task type are described (Section 2), with a focus on studies using MAIN. Then, the aim and research questions of the present study are stated (Section 3), followed by a description of the methodology (Section 4), and the results (Section 5). The final section of the paper (Section 6) contains the discussion and conclusion.

2. Literature review: age and task effects on story comprehension and story structure

In this section, an overview of previous studies investigating the effects of age and task type (retelling vs. telling) on children's story comprehension and story structure is given. First, in Section 2.1, general findings from studies that investigated task effects or the combination of age and task effects using various types of picture-based stimuli are briefly summarized; for reasons of space, studies that only investigate age effects are not included here. In Section 2.2, there is a detailed discussion of studies using MAIN, the stimuli used in the present study, including studies investigating age effects and development, as these results are directly comparable to the present study.

2.1. General findings

Studies using both telling and retelling tasks have often used different stimulus materials for the two tasks (e.g., Ripich and Griffith, 1988; Merritt and Liles, 1989; Gutiérrez-Clellen, 2002), or even different tasks for telling and retelling, e.g., retelling a story they heard and generating their own story from pictures (Isbell et al., 2004) or telling a personal narrative and retelling the contents of a storybook (Kaderavek and Sulzby, 2000), making it difficult or even impossible to compare the results from the two tasks; any difference could be linked to the differences in the stimuli and not to the task types. In some cases, retelling and telling were both investigated, but the results from the two tasks were not compared (Merritt and Liles, 1987; Reuterskiöld Wagner et al., 1999; Gutiérrez-Clellen, 2002; Walker et al., 2023). To my knowledge, no previous study has investigated how children's ability to answer comprehension questions about the story differs

¹ In MAIN, the task types are called *elicitation modes*. There is also a third task type, model story, where the child listens to the story and then answers the comprehension questions without having retold the story.

between telling and retelling tasks, using other stimuli than MAIN (for an overview of results from studies using MAIN; see section 2.2). However, results from the few studies investigating the effect of task type on measures of story structure (narrative macrostructure) using the same type of materials in both tasks indicate that performance is higher on retelling tasks (e.g., Schneider and Dubé, 2005; Lever and Sénéchal, 2011; see also section 2.2). Children as young as 5 years have been found to perform better on retelling tasks. For example, Lever and Sénéchal (2011) investigated telling and retelling by English-speaking mono- and bilingual 5-year-olds ($N = 50$) using the picture-based narrative tasks of the Edmonton Narrative Norms Instrument (ENNI; Schneider et al., 2005). While the primary aim of their paper was to investigate the effect of a dialogic reading training, they also compared the children's performance on story grammar units in telling and retelling (with pictures present in both tasks), and found significantly higher scores for the retelling task.

One previous study investigated the interaction between age and task type using the same stimuli for the different tasks. In their study of 44 English-speaking children in Kindergarten (aged 5–6) and Grade 2 (aged 7–8) using five pictures from a wordless book (and an accompanying story script) as stimuli, Schneider and Dubé (2005) compared three different tasks, one where the children only heard the story (“Oral only”) and then retold it, one where they only saw the pictures (“Pictures only”) and then told the story, and one where they both heard the story and saw the pictures (“Combined”) before they retold it. The order of the tasks was counterbalanced and all children did all tasks. They analyzed the children's performance on the production of story grammar units, a measure similar to the story structure score used in the present study, and found a significant effect of both age group and task type. However, the older group only performed significantly higher than the younger group on the “Oral only” and the “Combined” task. In the younger group, there was only a significant difference between “Combined” and “Pictures only” tasks, with higher scores on the former, whereas in the older group, performance was higher on both “Oral only” and “Combined” compared to “Pictures only.” There thus seems to be an interaction between age and task type, indicating that older children benefit more from hearing the story than younger children, and that differences between age groups depend on the task type. The present study investigates this issue further, by analyzing the interaction between task type and age group, for the first time in a study using MAIN with children above age 7.

2.2. Studies using MAIN

Only three previous studies using MAIN to investigate story comprehension and story structure have analyzed the interaction between age and task type (Roch et al., 2016; Wehmeier, 2019, 2020). For this reason, separate descriptions are given of studies investigating age effects (Section 2.2.1), either in telling or in retelling, and studies investigating the effect of task type (Section 2.2.2), i.e., the difference between telling and retelling; the three studies investigating the interaction effect are described in the latter section.

2.2.1. Age effects

Studies that investigated narratives elicited with MAIN have found clear age effects for both story comprehension and story structure in children aged 4–7, for a large number of different languages and language combinations, including Swedish (e.g., Bohnacker, 2016; Kunnari et al., 2016; Roch et al., 2016; Lindgren and Bohnacker, 2020; Wehmeier, 2020; Bohnacker et al., 2022, to name a few). Age effects were similar for studies using the telling task and those using the retelling task. In addition to results for the overall comprehension and story structure scores, some studies also report results for different comprehension questions (e.g., Bohnacker et al., 2020; Lindgren and Bohnacker, 2020) or types of story structure components (e.g., Lindgren, 2018; Öztekin, 2019), as is also done in the present study. For example, Lindgren and Bohnacker (2020) showed a significant age effect on German-Swedish bilingual children's overall performance on the comprehension questions, but the effect was larger for specific questions and performance on the questions differed. For example, the children scored substantially lower on the Cat/Dog question D10 compared to the other questions, and while the 6-year-olds only scored slightly higher than the 4-year-olds on D10, when answering the question incorrectly the older group gave reasonable explanations for their answer to a larger extent than the younger. Lindgren (2018) found for Swedish monolingual and German-Swedish and Turkish-Swedish bilingual 4 to 6-year-olds that an age effect was mainly seen for story structure components that were visible in the pictures (attempts, outcomes), and only to a smaller extent for those that have to be inferred (goals, internal states).

Only a smaller number of studies using MAIN included 8-year-olds or older children and their results are more mixed, both for story comprehension and story structure (e.g., Gagarina, 2016; Fiani et al., 2020, 2022; Košutar et al., 2022; Lindgren, 2022; Tribushinina et al., 2022). Most relevant here are studies investigating children aged 6 to 8, the age groups included in the present study. These studies all used a telling task, did not investigate specific comprehension questions or macrostructural components, and only one study of Swedish-speaking children included children above age 7 (Lindgren, 2022). In her longitudinal study of 17 Swedish-speaking children from age 4 to 9, Lindgren (2022) found significant development from age 7 to 9 in story structure, but not in story comprehension, due to the overall high scores for this measure, with scores above 9 points out of 10 already at age 5–6. However, in their study of Lebanese Arabic-French bilinguals ($N = 48$), Fiani et al. (2020) did find a significant effect of age on story comprehension from age 4 to 9, as did Peristeri et al. (2020) for Albanian-Greek bilinguals and Greek monolinguals aged 6 to 8. For story structure, three studies that investigated bilinguals did not find any age effect for children above age 6 (Gagarina, 2016; Fiani et al., 2022; Yang et al., 2023), whereas two studies of monolinguals did find an effect: the study by Lindgren (2022) described above, and a study of 89 Croatian-speaking monolinguals, where a significant difference was found between 6 and 8-year-olds (Košutar et al., 2022). It is possible that the lack of a significant age effect in the studies of bilinguals is linked to the more diverse language backgrounds of bilinguals, compared to monolinguals.

To summarize, studies using the MAIN measures of story comprehension and story structure found a consistent development during the preschool period, but not in the early school age. The present study adds to our knowledge of children's narrative abilities and how these are affected by age and task type by investigating story comprehension and story structure, including performance on different comprehension questions and macrostructural components in 6-year-olds and 8-year-olds using both a telling and retelling task.

2.2.2. Effects of task type: telling vs. retelling

Relatively many previous studies have investigated the effect of task type (telling vs. retelling) using MAIN, analyzing story comprehension (Kunnari and Välimaa, 2020; Wehmeier, 2020), story structure (Kunnari et al., 2016; Kuvač Kraljević et al., 2020; Sheng et al., 2020), or a combination of the two (Maviş et al., 2016; Roch et al., 2016; Wehmeier, 2019; Otwinowska et al., 2020). An overview of these studies can be found in the Appendix (Table A1). It should be noted that none of these studies included children as old as 8 years, and none of them investigated Swedish-speaking children. Before summarizing their results, it is necessary to point out that there are three methodological issues with these studies.

First, all studies followed the same procedure for the telling task (as described in Gagarina et al., 2019b), where a method of non-shared visual attention and non-shared knowledge is used. This means that, before and during the storytelling, the pictures are only visible to the child, and that the child chooses an envelope out of three, which are said to contain different pictures, creating a belief that the experimenter does not know which story the child is telling. However, in all studies cited above except two (Roch et al., 2016; Kuvač Kraljević et al., 2020), the method employed in retelling was different in terms of non-shared knowledge, as, after the child chose the envelope, the experimenter read the story; knowledge of the specific content of the (model) story was thus shared between experimenter and child, something which could influence the child's narration and which poses a problem when comparing results from the two tasks. With shared knowledge between child and experimenter, the situation may be more naturalistic compared to the situation of the telling task, but the child could also be less inclined to tell a detailed story; when shared knowledge is used in retelling but not in telling, this may thus either over- or underestimate the effect of task type. This is especially true when the pictures are visible to both child and experimenter during the reading of the story and the retelling, as was the case in the study by Otwinowska et al. (2020). The studies by Roch et al. (2016) and Kuvač Kraljević et al. (2020) instead used a method where the child listened to a pre-recorded story over headphones, which makes the retelling procedure more similar to the telling procedure in terms of shared knowledge. The same type of procedure is used in the present study. No previous study using MAIN has compared the effect of using a recorded story compared with a story read by the experimenters on children's narrative productions, but these differences in the methodologies should be kept in mind

when comparing the results of the present study with those of previous studies.²

Second, and most importantly, the studies published so far all used the Cat/Dog stories for retelling and the Baby Birds/Baby Goats stories for telling.³ This was the intention in the original version of MAIN (Gagarina et al., 2012), but was changed in the revised MAIN (Gagarina et al., 2019b), where instructions for both task types are available for all four stories. The four MAIN-stories, Cat, Dog, Baby Birds and Baby Goats were created to be parallel in their macrostructure. However, while it is true that the same number and types of macrostructural components are depicted in or can be inferred from all four sets of pictures, and that the comprehension questions are generally similar, there are some important differences between on one hand, Cat/Dog and, on the other, Baby Birds/Baby Goats. In Baby Birds/Baby Goats, events are mainly organized in a linear fashion, with one episode following the previous one, whereas in Cat/Dog, multiple events that belong to different episodes take place simultaneously. Telling Cat/Dog is thus more complex, as it requires the narrator to describe events that take place at the same time (see Kawar et al., 2023). There are also differences between the two pairs of stories in terms of their story characters. Whereas, Cat/Dog has three characters and includes one human (a boy), Baby Birds/Baby Goats has five characters, which are all animals. Here, Baby Birds/Baby Goats are thus more complex and may pose a greater challenge for the narrator. Using Cat/Dog for retelling and Baby Birds/Baby Goats for telling, as has been done in all previous studies, thus leads to a possible confound between the effect of stimuli and the effect of task type.

The above-mentioned differences between Cat/Dog and Baby Birds/Baby Goats seem more likely to influence the child's narrative production, and should thus pose a smaller problem for studies of story comprehension. However, studies that investigated story comprehension using a design where all children told one story from each pair (i.e., a within-subjects' design) with the same task type have found significant differences between Cat/Dog and Baby Birds/Baby Goats, with higher scores on the former compared to the latter (Bohnacker et al., 2020; Lindgren and Bohnacker, 2020; Bohnacker and Lindgren, 2021; Kawar et al., 2023).⁴ In most of these studies, the children told Cat/Dog first, at the beginning of a testing session containing multiple tasks, and Baby Birds/Baby Goats at the end, and it is thus possible that the significantly higher performance is caused by the children experiencing fatigue while answering the questions to Baby Birds/Baby Goats. However, the authors of these studies also point to important differences

² A previous study using another type of stimuli did find that typically-developing children produced narratives with more well-formed story structure when retelling the story to a naïve listener (Liles, 1987), indicating that there may indeed be a difference between hearing the story read by the experimenter and listening to it over headphones.

³ Wehmeier (2019, 2020) used the Cat story for retelling and the Baby Birds for telling.

⁴ In Lindgren and Bohnacker (2020), the difference was significant for the 4- and 5-year-olds but not for the 6-year-olds, which was likely due to the high scores on both tasks in this group.

between the story pairs in the level of difficulty of some of the comprehension questions (see Lindgren and Bohnacker, 2020; Bohnacker and Lindgren, 2021), which could explain the results. Children's expressions of story structure in narratives elicited with Cat/Dog and Baby Birds/Baby Goats, where Cat/Dog was also told first, have not been found to differ significantly (Lindgren, 2018; Bohnacker et al., 2022; Lindgren and Bohnacker, 2022). As these studies did not counterbalance the order of the two pairs, their results should be interpreted with caution, but they point toward differences between the two pairs, at least in comprehension. Studies that want to isolate the effect of task type thus need to control for the stories used in the two tasks. For this reason, the present study uses Cat/Dog in both telling and retelling.

Finally, only one previous study (Roch et al., 2016) counterbalanced the order of the retelling and telling tasks; in all other studies, the children received the tasks in a fixed order, either retelling-telling (Wehmeier, 2019, 2020; Sheng et al., 2020) or telling-retelling (Kunnari et al., 2016; Maviş et al., 2016; Kunnari and Välimaa, 2020; Kuváč Kraljević et al., 2020; Otwinowska et al., 2020). Any effect of task type found in studies where the order was not randomized could thus be an order effect. In the present study, order of the tasks was therefore counterbalanced within the age groups.

Let us now turn to the results from the studies of telling and retelling using MAIN, keeping their limitations in mind. For both story comprehension and story structure, some studies found a significant effect of task type, with better performance in retelling than in telling, whereas others found no significant difference between the tasks. Three studies found significantly higher scores on *story comprehension* after retelling than after telling (Roch et al., 2016; Otwinowska et al., 2020; Wehmeier, 2020), whereas two studies found no significant difference (Maviş et al., 2016; Kunnari and Välimaa, 2020). Here, it is worth mentioning that Roch et al. (2016), in the only study that controlled for the order of tasks and used a prerecorded story in the retelling task, showed a large difference in the scores of telling and retelling. For *story structure*, four studies found significantly higher scores in retelling (Roch et al., 2016; Wehmeier, 2019; Kuváč Kraljević et al., 2020; Otwinowska et al., 2020), whereas one found no significant difference (Maviş et al., 2016), and two found only partial differences: Kunnari et al. (2016), in their study of Swedish-Finnish bilinguals ($N = 16$) and Finnish monolinguals ($N = 16$), found a difference only for the bilinguals in Finnish, and Sheng et al. (2020) found significantly higher scores in retelling than in telling for children who were at risk for developmental language disorder (DLD), but not in typically-developing (TD) children. The studies that found no significant difference tended to have a lower number of participants than those who found an effect (e.g., $N = 13$ in Maviş et al., 2016 compared with $N = 62$ in Roch et al., 2016 and $N = 198$ in Wehmeier, 2019), which suggests that the former may have suffered from a lack of power. However, it is still possible that the task effect is caused by differences in the stories used in the two tasks or is connected to the order of the tasks.

Regarding the interaction between age and task type, two studies of German-speaking children aged 4;6–5;11 divided into three age groups (4;6–4;11, 5;0–5;5, 5;6–5;11) investigated the interaction between age and task type, one for story comprehension

of 199 monolingual and 66 bilingual children (Wehmeier, 2020), and one for story structure in 198 monolingual children (Wehmeier, 2019).⁵ Additionally, Roch et al. (2016) investigated the interaction for both story comprehension and story structure in a study of Italian-English bilingual 5–7-year-olds ($N = 62$), divided into two age groups (5–6, 6–7). None of the studies found a significant interaction effect, which may be explained by the fact that the groups were very close to each other in age; in fact, analyzing the effect of age as a linear variable would have been more appropriate in these studies.

In sum, there are indications that task type influences story structure and story comprehension also when MAIN is used, but results are somewhat mixed and should be interpreted with caution since the studies all used different stories for the two tasks, with two exceptions, used a method with shared knowledge between child and experimenter in retelling, and, with one exception, did not control for the order of the tasks. Studies wanting to isolate the effects of task type should thus control for the stories used in the different tasks, the amount of shared knowledge between experimenter and child, and the order of the tasks, as is done in the present study.

3. The present study

The present study investigates the effect of task type (retelling vs. telling) in Swedish-speaking children from two different age groups (6-year-olds vs. 8-year-olds) on story comprehension and story structure (narrative macrostructure). The following research questions are asked:

- How do task type and age affect story comprehension and story structure?
- Is the effect of task type the same in both age groups?
- Is the difference between the age groups the same for both tasks?

Based on results from the previous studies cited above, it is predicted that performance will generally be higher in retelling than in telling, as having heard a model story will make it easier for the children to provide correct answers to comprehension questions, and to produce the expected macrostructural components. Story comprehension will be very high already at age 6, possibly close to ceiling, but it is expected that there will be some effect of task type, with better comprehension in retelling. It is also expected that the 8-year-olds will perform better on both tasks than the 6-year-olds. However, these general effects will be modulated by an interaction effect, with a smaller difference between the age groups in telling and thus a stronger effect of task type in the older children.

⁵ The participants in these two studies are partially the same and Wehmeier (2019) also reports results for story comprehension for 198 out of the 199 monolingual children included in Wehmeier (2020).

4. Methods

4.1. Participants

The data collection took place from January to March 2022. The study was approved by the Swedish Ethical Review Authority [No: 2021-06403-01]. Seventy-four children from a school in a larger city in central Sweden participated in the study, 36 8-year-olds (16 girls), who attended second grade (mean age: 8;6, range: 8;2–9;1), and 38 6-year-olds (17 girls), who attended *förskoleklass* “preschool class,” a preparatory year before first grade (mean age: 6;8, range: 6;3–7;2). There were three pairs of siblings. The children came from homes with mid- to high SES, as measured by parental education. All parents had completed secondary education, and the majority of the children (88%, 65 children) had two parents who had at least completed a bachelor’s degree. All children had Swedish as their first language: 13 children (7 6-year-olds, 6 8-year-olds) grew up bilingually, hearing both Swedish and another language at home (4 of these children were reported to only understand or speak these languages to a limited extent), and the remaining 61 children grew up in homes where only Swedish was spoken. All children had at least one parent who was a speaker of Swedish as a first language. The other languages spoken in the homes were English, Estonian, Farsi, Finnish, German, Italian, Kiswahili, Russian, Slovak and Spanish. Additionally, six children, three in each age group, were reported to speak English, where English was not a language spoken in their homes. All children except two were born in Sweden. These two children had arrived in Sweden at the ages of 1 and 2 months, respectively. According to parental report, the children all had typical language development.

4.2. Materials and procedure

All participants carried out two picture-based narrative tasks from MAIN, the *Cat* and the *Dog* story (Gagarina et al., 2019b).⁶ The *Cat* story and the *Dog* story each consist of six colored pictures and a story script. The stories have been carefully created to be parallel in their characters (two animals, one boy) and the content and structure of their three episodes, including the macrostructural components that are depicted in the pictures and are verbalized in the story scripts. MAIN contains scoring protocols for standardized measures of story comprehension and story structure (narrative macrostructure), which were both analyzed in the present study (see Section Coding and analyses).⁷ The model used to assess *story structure* in MAIN is based on goal-based story grammar models (e.g., Mandler and Johnson, 1977; Stein and Glenn, 1979; Stein and Policastro, 1984), and on the decision-tree model of narrative

structure by Westby (2012). In goal-based models, narratives consist of a setting (time and place of the events) and at least one episode, and the goal is seen as the core of the episode, the component around which events are organized. In the MAIN model, each episode contains five different types of macrostructural components: internal state as initiating events, goal, attempt, outcome, and internal state as reaction. The picture sequences used as stimuli as well as the story scripts include these components. An overview of the different types of components is given in the Appendix (Table A2). Table 1 gives a summary of the content of the episodes of the two stories.

Story comprehension in MAIN is assessed with 10 comprehension questions, that measure the child’s ability to draw inferences with respect to the character’s internal (mental) states and the reasons for their actions (goals) as well as their understanding of the plotline as a whole. The comprehension questions of the *Cat* story can be found in the Appendix (Table A3).⁸

The children carried out the narrative tasks in a quiet room at the school as part of a session that included a number of additional tasks. The author administered the tasks to all children. Each child received one story as a *retelling* task and the other one as a *telling* task (Gagarina et al., 2019b), where one task was administered at the beginning of the session and one toward the end. The order of the tasks was counterbalanced across participants within age groups, and half of the participants received the *Cat* story as a telling task and the *Dog* story as a retelling task and vice versa. The stories and the orders of the tasks were randomly assigned.

The MAIN standardized procedures were used (Gagarina et al., 2019b). For both telling and retelling, the child and the experimenter sit at opposite sides of a table on which three envelopes, each containing a set of story pictures, are placed. The child receives the information that the envelopes contain different stories, chooses an envelope, opens it and looks at the pictures, which are presented as one long strip. When the child has looked at all pictures, they are folded back so that only the first two are visible. In the telling task, the child then begins to tell the story. The next two pictures are unfolded when the child has finished telling the events in the first two pictures. Finally, all six pictures are visible. The experimenter is only allowed to give general prompts (e.g., *aha*, *mhm*, or *and then?*). When the child has finished his/her story, the experimenter asks the comprehension questions. During the questions, the pictures are on the table, visible to both experimenter and child. At no time before or during the story telling are the pictures visible to the experimenter.

In the retelling task, the child listens to the story before telling it. In the present study, the story was prerecorded by a female native speaker of Swedish, instead of read by the experimenter as in the original MAIN procedure. The retelling procedure was therefore as follows: the child is shown three pictures on a tablet and is asked to indicate which picture belongs to the story s/he has chosen (to give the impression that the experimenter does not

6 The revised Swedish version of MAIN, translated and adapted by Ute Bohnacker, was used.

7 The complete MAIN, including the picture sequences, instructions on how to administrate it, the comprehension questions and scoring guidelines for story comprehension and story structure, is freely available in a large number of different languages (e.g., Dutch, English, Spanish, Swedish, Turkish) from <https://main.leibniz-zas.de/en/main-materials/main-materials/> (registration is required to download the materials).

8 The comprehension questions for the *Dog* story only differ from those of the *Cat* story in the characters and objects included (e.g., dog instead of cat, mouse instead of butterfly, sausages instead of fish); as shown in Table 1, the plotlines are otherwise identical.

TABLE 1 Content of the three episodes of the *Cat* and the *Dog* story.

	Cat story	Dog story
Episode 1	A cat wants to catch a butterfly, tries, but falls into a bush and hurts itself.	A dog wants to catch a mouse, tries, but hits its head against a tree and hurts itself.
Episode 2	A boy loses his ball in the water, wants to get it back, uses his fishing rod, gets it back and is happy.	A boy loses his balloon in a tree, wants to get it back, jumps up, manages to get it back and is happy.
Episode 3	The cat sees fish that the boy brought, wants to eat them, takes them and eats them and is happy.	The dog sees sausages that the boy brought, wants to eat them, takes them and eats them and is happy.

know which story the child has chosen), and then s/he listens to the recording of the story over headphones before telling it and answering the comprehension questions, following the procedure described above. Using a prerecorded story instead of having the experimenter read it ensures that all children hear the story told in an identical way, and also that there is no shared knowledge of the story between experimenter and child.

4.3. Coding and analyses

The author transcribed all narratives orthographically using the CHAT-format (MacWhinney, 2000) and carried out all coding and analyses. Story comprehension and the production of story structure was analyzed using standardized measures from MAIN.

4.3.1. Story structure

Following the MAIN standardized scoring scheme (Gagarina et al., 2019b), points were awarded to each narrative for the successful production of setting (time + place, 2 points, e.g., *Once upon a time, next to a lake...*) and for internal states as initiating event (e.g., *A cat saw a butterfly*), goal (e.g., *The cat wanted to catch the butterfly*), attempt (e.g., *The cat jumped toward the butterfly*), outcome (e.g., *The cat got stuck in the bush*) and internal state as reaction (e.g., *The cat was angry*) in each of the three episodes (5 points x 3 episodes). This resulted in a story structure score (max = 17 points) for each narrative. The effects of task type (telling, retelling), age group and the interaction between task type and age group on this total score was analyzed with a linear mixed effects model using the *lmer*-function of the *lme4*-package (Bates et al., 2015) in R. To ensure that the story and the task order did not influence results, these were included as control variables in the model. Participant was included as random effect.⁹ Additionally, a closer inspection of the proportion successful verbalizations of the six different types of macrostructural components (setting, internal state as initiating event, goal, attempt, outcome, internal state as reaction) was carried out to find patterns related to age and task type.

4.3.2. Story comprehension

Ten standardized comprehension questions (Table A3) were asked after each narrative task. The questions are the same for both

stories; the only differences concern the characters/objects, which are different in the two stories. Three questions probe the child's understanding of the character's goals (e.g., *Why does the cat grab the fish?*), and three questions target internal states (e.g., *How does the dog feel?*), with three further questions seeking the rationale for the character's internal state (e.g., *Why does the dog feel bad?*). The final question addresses understanding of the overall plotline (e.g., *Will the boy be friends with the cat? Why?*). The answers to the comprehension questions were scored following the MAIN manual (Gagarina et al., 2019b). Each child received a story comprehension score for each task (Max = 10 points). Just as for story structure, a linear mixed effects model was run to investigate the effects of task type (telling, retelling), age group and the interaction between task type and age group, on the story comprehension score. Story and task order were included as control variables in the model. Participant was included as random effect. Additionally, a closer inspection of the accuracy for each of the comprehension questions was carried out and answers to questions with especially low scores were analyzed qualitatively.

5. Results

5.1. Story structure

Figure 1 shows the results for the story structure score by task type (retelling, telling) for the two age groups. The story structure scores are not close to ceiling; on average, the children produced only 49–65% out of the macrostructural components. Scores on the retelling task are on average around 2 points higher than scores on the telling task, and the mean scores of the 8-year-olds are 1 point higher than the scores of the 6-year-olds. The amount of individual variation, as measured by the standard deviations, is similar across age groups and tasks. The statistical analysis (Table 2) shows that there are significant effects of both age group and task type, with a larger effect size (β -value) for the latter than the former. The interaction effect was not significant, showing that both age groups are affected to the same extent by the task type. Task order, which was a control predictor, was significant; the children performed significantly better on the second than on the first task. The story (Cat, Dog) did not have a significant effect on the scores.

Age group and task type were thus both found to significantly affect the story structure score. The question is whether these effects are similar across the different types of macrostructural components or whether the effects are more prominent for some of the components. For each task, each child had the possibility to produce a maximum of two setting components (time, place), and

⁹ Since each participant had only two scores, one for telling and one for retelling, it was not possible to include random slopes in the models.

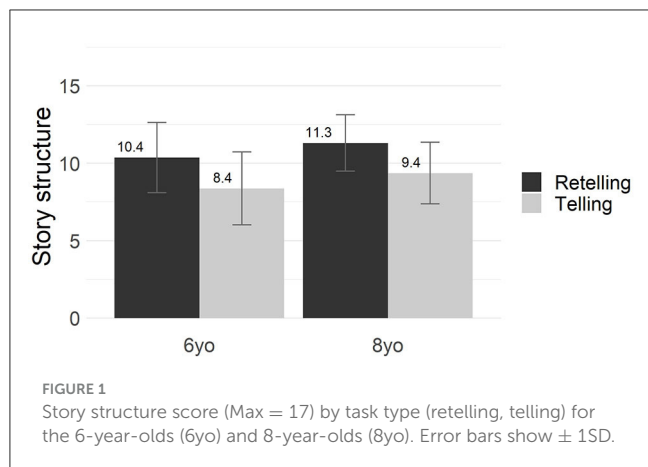


FIGURE 1
Story structure score (Max = 17) by task type (retelling, telling) for the 6-year-olds (6yo) and 8-year-olds (8yo). Error bars show \pm 1SD.

TABLE 2 Linear mixed effects model for the story structure score (Story structure \sim Age group * Task type + Task order + Story + 1 | Participant).

Random effects	Variance	SD			
Participant: intercept	2.42	1.56			
Fixed effects	β	SE	df	t	p
Intercept	10.01	0.38	134.08	26.62	<0.001***
Age group: 8yo vs. 6yo	0.99	0.49	109.59	2.04	0.04*
Task type: telling vs. retelling	-1.95	0.32	70.00	-6.12	<0.001***
Age group * task type	-0.05	0.46	70.00	-0.10	0.92
Task order: 2nd vs. 1st	0.86	0.23	70.00	3.75	<0.001***
Story: dog vs. cat	-0.20	0.23	70.00	-0.87	0.39

The model has been fit by REML. The t-tests use Satterthwaite's method. Values have been rounded to two decimal points. For each predictor, the second category is the reference level. * $p < 0.05$, *** $p < 0.001$.

a maximum of three internal states as initiating events (IS as IE), goals, attempts, outcomes and internal states as reaction (IS as R), one in each of the three episodes. Figure 2 shows the percentage of accurately verbalized components by task type for the two age groups for each of the six components.

The first observation is that there is a large difference in the extent to which different types of components are accurately verbalized by the children. Whereas, only around a third or less of the goals are produced, attempts and outcomes are almost always produced, irrespective of age group or task type. With the exception of goals in the telling task, the 8-year-olds produce all types of components more frequently than the 6-year-olds in both tasks. The age effect that was found in the statistical model (Table 2) is thus similar across component types; it is a general age effect, influencing the children's overall narrative skills and is not linked to development in the ability to produce a specific type of component. In contrast, the effect of task type is mainly found for internal states as initiating events (IS as IE) and as reaction (IS as R), and to a smaller degree also for goals. In retelling, the 8-year-olds produce 52% of all IS as R, compared to only 23% in telling; the corresponding figures for the 6-year-olds are 48% and 16%. In retelling, the children's production of IS as IE (6-year-olds: 69%, 8-year-olds: 79%) is close to the level of the attempts (6-year-olds: 69%, 8-year-olds: 90%), whereas in telling, it is substantially lower

(6-year-olds: 47%, 8-year-olds: 53%). The children's performance on the components that are not clearly visible in the pictures, but require inferencing, is thus boosted by hearing a model story, whereas the other components are at the same level in both task types.

5.2. Story comprehension

Figure 3 shows the results for story comprehension by task type (retelling, telling) for the two age groups. The scores are uniformly high, with means of around nine points across tasks in both groups and with small differences between the age groups and tasks. The statistical analysis shows that none of the predictors had any significant effect (Table 3).

Although the scores are generally close to ceiling (above 90% accuracy for the ten questions combined), the majority of the children do not score 10 points. The proportion of the children who score at maximum is not influenced by task type and is slightly lower for the 8-year-olds (6-year-olds telling: 45%, 17/38; 6-year-olds retelling: 42%, 16/38; 8-year-olds telling: 36%, 13/36; 8-year-olds retelling: 36%, 13/36). The majority of the 8-year-olds did thus not answer all questions correctly after hearing a model story. Why is this the case? Since previous studies have found that performance differs on the individual comprehension questions (e.g., Lindgren and Bohnacker, 2020), it is relevant to analyze accuracy for the ten comprehension questions separately to see if this may provide an answer. The results are shown in Figure 4.

As seen in Figure 4, accuracy is very high for the first nine questions (81.5% or higher), and for these questions, the accuracy is somewhat higher for the 8-year-olds and in retelling, as expected based on their slightly higher overall scores (see Figure 3). However, for D10, a question that investigates the child's understanding of the overall plotline, the performance is radically lower, with accuracy for the 8-year-olds below 50%, in both telling and retelling. The highest accuracy on this question is found for telling in the 6-year-old group (63%). If we look at the performance of individual children on this question, the majority of the children consistently answer the question either correctly or incorrectly across the two tasks (6-year-olds: 84%, 32/38; 8-year-olds: 75%, 27/36). No pattern can be found with respect to the order of the tasks and the task type for the children who answer one question correctly and one incorrectly, i.e., it is not always the case that the correct answer is for the telling or retelling task or for the first or the second task. Having heard the oral model story does thus not help children answer this question correctly and having already performed a similar narrative task also does not help. Is there any explanation for this finding? To answer this question, it is necessary to take a closer look at the children's incorrect answers.

The question D10 is *Will the boy be friends with the cat/dog? Why?* The expected answer is either *No, because the cat/dog ate the boys' fish/sausages* or *Yes, if the boy forgives the cat/dog (for eating the fish/sausages)* (or other formulations with the same meaning). This type of answer is the only answer that is scored as correct following the MAIN manual. The children's incorrect answers were coded following the categories

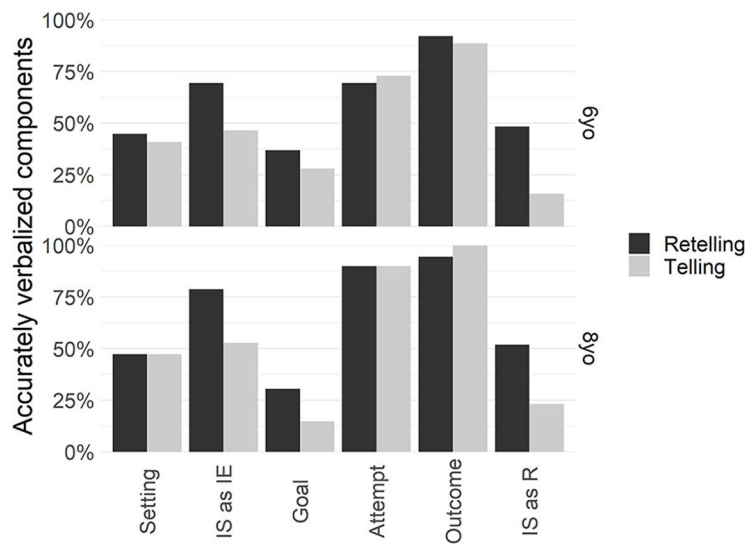


FIGURE 2 The percentage accurately verbalized components by task type (retelling, telling) for the 6-year-olds (6yo) and 8-year-olds (8yo). (IS as IE = internal state as initiating event; IS as R = internal state as reaction).

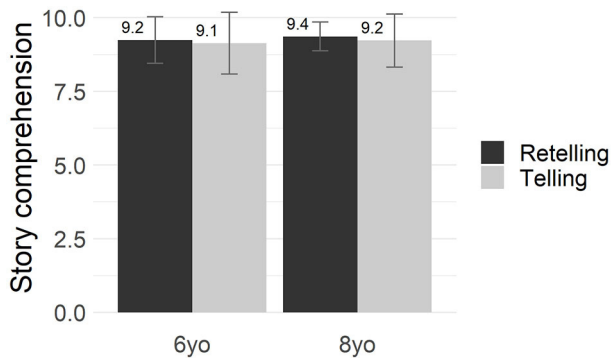


FIGURE 3 Story comprehension (Max = 10) by task type (retelling, telling) for the 6-year-olds (6yo) and 8-year-olds (8yo). Error bars show $\pm 1SD$.

TABLE 3 Linear mixed effects model for the story comprehension score (Story comprehension \sim Age group * Task type + Task order + Story + (1 | Participant)).

Random effects	Variance	SD			
Participant: intercept	0.24	0.49			
Fixed effects	β	SE	df	t	p
Intercept	9.26	0.16	141.95	59.11	<0.001***
Age group: 8yo vs. 6yo	0.14	0.19	127.81	0.70	0.47
Task type: telling vs. retelling	-0.09	0.15	70.00	-0.59	0.56
Age group * Task type	-0.06	0.22	70.00	-0.25	0.81
Task order: 2nd vs. 1st	0.11	0.11	70.00	1.03	0.31
Story: dog vs. cat	-0.17	0.11	70.00	-1.52	0.13

The model has been fit by REML. The t-tests use Satterthwaite’s method. For each predictor, the second category is the reference level. Values have been rounded to two decimal points. ***p < 0.001.

established by Lindgren and Bohnacker (2020). Table 4 shows the types of answers for the two age groups. No pattern could be found with respect to the types of incorrect answers and task type; for this reason, the table shows the number of incorrect answers for both task types combined. The specific story (Cat or Dog) was not linked to the type of incorrect answer either; incorrect answers were equally frequent with both of the two stories.

Table 4 shows that the children give different types of incorrect answers. Some children state that they do not know the answer. A few cryptic or nonsensical answers, such as *Kanske, för katter är katter* “Maybe, because cats are cats” (2K-07, telling), are given. A number of answers show that the child thinks that the question is about whether the cat/dog will be friends with the boy instead of the other way around, e.g., *Kanske, för att hunden fick pojakens korvar så blev hunden kompis med honom* “Maybe, since the dog

got the boy’s sausage then the dog became friends with him” (FK-01, retelling); this category was added to the classification by Lindgren and Bohnacker (2020). The remaining incorrect answers are all reasonable explanations, based on world knowledge, especially of relationships between cats/dogs, which typically are pets in the cultural context where these children grow up, and people. Examples of such answers are *Tror det, för att hundar och människor brukar komma rätt bra överens* “[I] think so, because dogs and people usually get along pretty well” (2K-13, retelling), *Det vet jag väl inte, nej, för att det verkar som att hunden redan har en ägare* “That I do not know, no, because it seems as if the dog already has an owner” (2K-31, retelling), and *Aa, för att han tycker om hundar* “Yeah, because he likes dogs” (2K-40, telling). Two 8-year-olds also point out that it is not possible to know the answer, e.g., *Kanske, därför man ser inget svar* “Maybe, because one does not see an answer” (2K-14, telling), alluding to the fact that

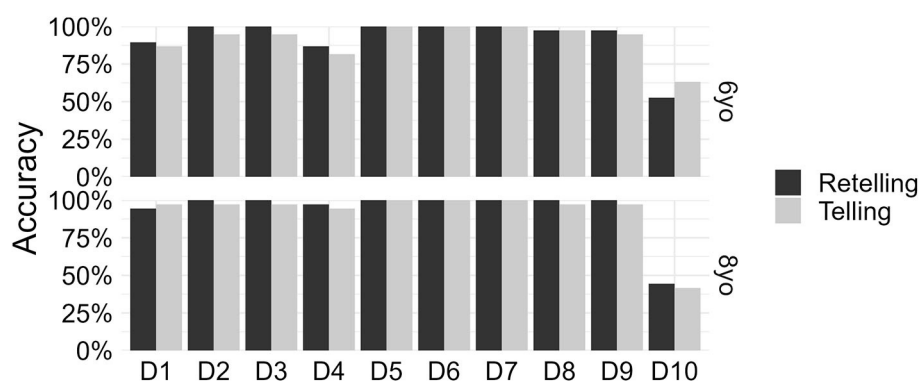


FIGURE 4

Accuracy for the ten comprehension questions by task type (retelling, telling) for the 6-year-olds (6yo) and 8-year-olds (8yo).

TABLE 4 Types of answers, question D10 (Will the boy be friends with the cat/dog? Why?), for the 6-year-olds (6yo) and 8-year-olds (8yo).

	6yo	8yo
Correct answers	44	31
Incorrect answers	32	41
I don't know/no explanation	7	7
Cryptic/nonsensical answer	2	2
Yes, the cat/dog will be friends with the boy	7	7
Reasonable explanations	16	25
Total number of answers	76	72

it is not shown in the pictures what would happen next. In the 8-year-old group, six answers state that it is typical for characters to become friends at the end of fairytales, e.g., *Tror det, för så är det i dom flesta sagor* “[I] think so, because that is how it is in most fairytales” (2K-35, telling), showing an understanding of the genre. The majority of the children who answered the question incorrectly on both tasks gave the same type of incorrect answer. Overall, these types of reasonable explanations constitute 50% (16/32) of the 6-year-olds’ and 61% (25/41) of the 8-year-olds’ incorrect answers. There is thus a tendency for the older children to give a higher percentage of reasonable, but incorrect answers. We return to these results in the discussion below.

6. Discussion and conclusion

The present study investigated the effects of task type (retelling vs. telling) and age (6-year-olds vs. 8-year-olds), as well as the interaction between task type and age, on 74 Swedish-speaking children’s story comprehension and production of story structure (narrative macrostructure). The Cat and Dog picture-based stories of the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2019b) were used to elicit two narratives and two sets of answers to the accompanying comprehension questions from each child, once as a telling task and once as a retelling task. Linear mixed effects models were used to analyze the effects of task

type and age and their interaction on the story comprehension and story structure scores. Task order and story were controlled for. It was predicted that performance would be higher in retelling than in telling and that the 8-year-olds would score higher than the 6-year-olds, but that there would also be an interaction effect, with a stronger effect of task type for the older children. The predictions were partially born out.

For *story structure*, as predicted, the analysis showed significant effects of both age and task type. The results for story structure with respect to the main effect of age are in line with previous studies investigating MAIN story structure in monolingual children of similar ages using a telling task (Košutar et al., 2022; Lindgren, 2022). Based on the present study, it can thus be concluded that, for both telling and retelling, children’s ability to adequately express story structure develops from the preschool to the early school age. Additionally, the scores do not show complete mastery of all aspects of story structure (no ceiling effect) even for the 8-year-olds in retelling, and the 8-year-olds’ scores in telling are lower than those found for Swedish-speaking adults (Gagarina et al., 2019a), meaning that the MAIN story structure score can show development of narrative skills beyond age 8, both in telling and in retelling.

The present study also found a significant task effect, with higher scores in retelling than in telling. This result is similar to findings from some of the previous studies (Roch et al., 2016; Wehmeier, 2019; Kuvač Kraljević et al., 2020; Otwinowska et al., 2020), but here found for the first time in a study controlling for task order, shared visual attention in retelling and the stories used. Importantly, the present study found a significant effect of task order, with higher scores on the second task than on the first one. Counterbalancing the order of the telling and retelling tasks across participants, which has only been done in one previous study using MAIN (Roch et al., 2016), is thus important, as the order of the tasks influences the children’s results. Although the results of some studies that did not counterbalance the order of the tasks were in line with those of the present study, the size of the differences between telling and retelling may still have been influenced by the task order, leading to either over- or underestimations of the true task effect. The significant effect of task order also shows that 6-year-olds and 8-year-old children’s ability to express story

structure improves with increased exposure to narrative tasks, a finding which has implications for efforts to improve children's narrative skills, e.g., within school settings. This finding indicates that primary school children's ability to express story structure can be significantly improved, when they first receive another narrative task, at least in the short term and when the tasks are highly similar. Future studies should investigate whether this is also the case when less similar tasks are used.

Contrary to the predictions, the interaction between age and task type was not significant. Both groups of children were affected by the task type to the same extent, on average scoring around 2 points higher in retelling than in telling. The task effect was stronger than the effect of age group, with the 8-year-olds scoring on average around one point higher than the 6-year-olds across tasks. The lack of a significant interaction effect was in contrast to the results from [Schneider and Dubé \(2005\)](#), who, using another set of stimuli than MAIN, found that the performance of 6-year-olds and 8-year-olds did not differ on a telling task, but that the 8-year-olds performed significantly better on retelling tasks. However, the results of the present study were in line with the results from two studies using MAIN, [Wehmeier \(2019\)](#) of children aged 4;6–6;11, and [Roch et al. \(2016\)](#) of children aged 5–7. It is possible that the difference in results between the present study and the study by [Schneider and Dubé \(2005\)](#) is linked to differences in the stimuli, for example in the number of characters (two in the stimuli used by [Schneider and Dubé, 2005](#), three in the stimuli of the present study), the number of episodes (one in [Schneider and Dubé](#), three in the present study), or in the scoring scheme used (differences in the types of macrostructural components scored, e.g., that the scoring scheme of [Schneider and Dubé, 2005](#), in contrast to the present study, awards points for character introductions, as well as initiating events). For example, it is possible that the older children in [Schneider and Dubé \(2005\)](#) scored higher in retelling compared to the younger children specifically on component types that are not included in the MAIN story structure score, or that with a shorter story with a simpler episodic structure and fewer characters, the ability to retell the story was more linked to memory, which is likely more developed in older children. Future studies should compare the same children's performance across different narrative stimuli, both in telling and in retelling, and investigate both younger and older ages, to enable more general conclusions about interactions between task type, stimuli and age.

In addition to the overall story structure score, the present study also included an analysis of the extent to which different types of macrostructural components were produced. The results showed an age effect across component types. This means that a general development in children's narrative ability takes place from age 6 to age 8, and not only a development of a more specialized ability that primarily influences the production of specific components. This result differed from previous studies that investigated types of components in 4–6/7-year-olds ([Lindgren, 2018](#); [Öztekın, 2019](#)), where development was mainly seen in the components that are visible in the pictures (attempts, outcomes) and less in those that need to be inferred (goals, internal states). The reason for this difference may be that those components that need to be inferred are infrequently produced at younger ages, and that development of these components primarily takes place after age 6. The present

study offers some support for this idea, as for example only around a third or less of goals were produced whereas attempts and outcomes were almost always (>80%) produced by both age groups. In contrast to this general effect of age, the task effect was mainly found for internal states, both as initiating events and as reactions, and to a smaller extent for goals. This may partly be due to the already high level of production for attempts and outcomes, but it is noteworthy that the children's performance on components that require inferencing are boosted by hearing a model story. In their tellings, the children rarely verbalized these inferred components, which are central to the production of a well-formed and expressive narrative, but they frequently included them in their retellings. This result has important implications for the possibilities to train children in the inclusion of such inferred components, e.g., as part of the school curriculum. The results from the present study indicate that a simple exposure to a model story leads to substantial improvement in the children's story structure, especially for the inferred components. It is likely that an intensive training scheme, where various types of model stories are used, and where the child repeatedly carries out different types of narrative tasks, would lead to even larger gains. In the future, studies should investigate whether first listening to a model story and then telling another story leads to a similar increase in performance, as well as whether improvement can be seen if the story is retold later, e.g., after 1 day or 1 week. Such studies would make it possible to find out whether the effect of hearing a model story can generalize and remains over time, allowing for stronger claims as to the positive effects of exposure to model stories on children's narrative skills. Future studies should also investigate whether younger children's ability to produce inferred components can also be similarly enhanced by hearing a model story.

Most previous studies using MAIN employed a method where there was a difference between the telling and retelling tasks in the shared knowledge between the experimenter and the child (see Section Effects of task type: telling vs. retelling), whereas the present study employed a method of non-shared knowledge for both tasks. However, results for story structure from the present study are similar to other studies that used a procedure with shared knowledge in the retelling task. It is possible that this is linked to the specific narrative aspect analyzed; it could be the case that the child's ability to express story structure is not influenced by shared knowledge with the experimenter in the same way as other narrative aspects such as the ability to introduce story characters appropriately (e.g., [Kail and Hickmann, 1992](#)). However, controlling for the procedure used is still important to ensure that the two tasks are as comparable as possible and that results are due to a difference between telling and retelling and not to other factors, such as the amount of shared knowledge, task order or the stimuli used. Here, the present study has offered a valuable contribution to studies of children's narrative skills. Future studies should investigate other narrative measures using the same controlled methodology.

For *story comprehension*, contrary to the predictions, the results showed no effect of task type or age, nor an interaction between these two factors. The overall accuracy was high (>90%) across tasks and age groups, which likely explains the result. This result differed from previous studies, possibly due to the stories used and

the age groups investigated. Regarding the ages investigated, as shown in both the present study and in previous studies (see e.g., Bohnacker and Lindgren, 2021; Lindgren, 2022), children aged 6 receive high scores on the Cat/Dog comprehension questions in telling, which does not leave a lot of room for higher scores on the retelling task, nor for an age effect. With respect to the stories, previous studies with MAIN that compared telling and retelling consistently used Cat/Dog in retelling and Baby Birds/Baby Goats in telling. Since studies that employed both sets of stories in the telling mode have shown significantly higher scores for Cat/Dog (Bohnacker et al., 2020; Lindgren and Bohnacker, 2020; Bohnacker and Lindgren, 2021), it is possible that the previously found effect of task type on story comprehension (e.g., Roch et al., 2016) is an effect of the stories used and not of task type. The retelling story may have been easier than the telling story. This shows the importance of carefully controlling the stimuli used in different types of tasks. To enable firm conclusions about the effect of task type on story comprehension, future studies should investigate younger children's story comprehension using the same controlled method as in the present study.

Although performance on the Cat/Dog comprehension questions has previously been found to be high already at age 5 (e.g., Bohnacker and Lindgren, 2021; Lindgren, 2022), the majority of the 8-year-olds were not able to answer question D10 (*Will the boy be friends with the cat/dog? Why?*) correctly after hearing a model story. For this specific question, which aims to assess the ability to understand the entire plotline, the 8-year-olds even had lower accuracy than the 6-year-olds. However, the 8-year-olds more often gave reasonable answers that were based on accurate world knowledge, e.g., that cats/dogs are pets and that therefore the boy would be expected to be friends with the cat/dog or that characters tend to become friends at the end of fairy tales, but that were not the expected correct answers (following the MAIN manual), showing that there is some development with age that cannot be seen in a purely quantitative analysis. The development can best be described as a general inferential ability that includes making use of other types of knowledge than those gained from the specific narrative task. This result is in line with the results from Lindgren and Bohnacker (2020), who found a similar qualitative development with age from age 4–5 to age 6 in German-Swedish bilinguals, where “[t]he 6-year-olds’ incorrect answers had a reasonable, logical explanation, whereas the younger children more often answered with a cryptic, seemingly nonsensical explanation or gave no explanation at all” (p. 92). Based on results from the present study, as well as others that have investigated performance on individual MAIN comprehension questions (Bohnacker et al., 2020; Lindgren and Bohnacker, 2020), it is clear that this development with age in children's general inferential ability does not lead to increased accuracy on the D10 question. On the contrary, giving these types of answers showing advanced reasoning is punished, whereas a “simpler” correct answer, based purely on the events of the story, is rewarded. With increasing age, children do not become more likely to offer the “correct” answers, but instead more frequently draw on their general inferential ability. In light of these results, I recommend an adjustment to MAIN so that these types of reasonable explanations are added as

examples of correct answers to the Cat/Dog question D10. Future studies should investigate the effect of such a change in the scoring procedure on the scores of children of different ages.

To conclude, the results of the present study confirm previous findings that expressing story structure in a retelling task is easier than in a telling task, but showing this for the first time while controlling for task order and stimulus complexity in MAIN. In addition, this study has extended previous work in including a group of 8-year-old children, who were shown to exhibit the same performance gap between telling and retelling as younger children, indicating that narrative abilities are still under development at this age.

Data availability statement

The datasets presented in this study and the scripts used to analyze the data are available in the OSF repository under the following link: <https://osf.io/j7y2u/>.

Ethics statement

The study was approved by the Swedish Ethical Review Authority [No: 2021-06403-01]. The study was conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians.

Author contributions

JL designed the study, collected and analyzed the data, and wrote the manuscript.

Acknowledgments

I thank the children and their parents for taking part in this study, and the teachers for their assistance with recruitment.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bohnacker, U. (2016). Tell me a story in English or Swedish: narrative production and comprehension in bilingual preschoolers and first graders. *Appl. Psycholinguist.* 37, 19–48. doi: 10.1017/S0142716415000405
- Bohnacker, U., and Lindgren, J. (2021). “MAIN story comprehension: what can we expect of a typically developing child?,” in *Language Impairment in Multilingual Settings: LITMUS in Action Across Europe*, eds S. Armon-Lotem, and K. K. Grohmann (Amsterdam: John Benjamins), 14–46. doi: 10.1075/tilar.29.01boh
- Bohnacker, U., Lindgren, J., and Öztekin, B. (2020). “Bilingual Turkish-Swedish children’s understanding of MAIN picture sequences: individual variation, age, language and task effects,” in *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives*, eds U. Bohnacker, and N. Gagarina (Amsterdam: John Benjamins), 99–147. doi: 10.1075/sibil.61.04boh
- Bohnacker, U., Lindgren, J., and Öztekin, B. (2022). Storytelling in bilingual Turkish-Swedish children: effects of language, age and exposure on narrative macrostructure. *Linguist. Approaches Biling.* 12, 413–445. doi: 10.1075/lab.20057.boh
- Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Lang. Teach. Ther.* 18, 1–21. doi: 10.1191/0265659002ct224oa
- Fiani, R., Henry, G., and Prévost, P. (2020). “Narrative comprehension in Lebanese Arabic-French bilingual children,” in *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives*, eds U. Bohnacker, and N. Gagarina (Amsterdam: John Benjamins), 32–60. doi: 10.1075/sibil.61.02fia
- Fiani, R., Henry, G., and Prévost, P. (2022). Macrostructure in narratives produced by Lebanese Arabic-French bilingual children: developmental trends and links with language dominance, exposure to narratives and lexical skills. *Linguist. Approaches Biling.* 12, 446–478. doi: 10.1075/lab.20035.fia
- Gagarina, N. (2016). Narratives of Russian-German preschool and primary school bilinguals: rasskaz and Erzählung. *Appl. Psycholinguist.* 37, 91–122. doi: 10.1017/S0142716415000430
- Gagarina, N., Bohnacker, U., and Lindgren, J. (2019a). Macrostructural organization of adults’ oral narrative texts. *ZAS Papers in Linguistics* 62, 190–208. doi: 10.21248/zaspil.62.2019.449
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balciuniene, I., et al. (2012). MAIN: multilingual assessment instrument for narratives. *ZAS Papers in Linguistics* 56, 1–140. doi: 10.21248/zaspil.56.2019.414
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balciuniene, I., et al. (2015). “Assessment of narrative abilities in bilingual children,” in *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 243–269. doi: 10.21832/9781783093137-011
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., et al. (2019b). MAIN: multilingual assessment instrument for narratives-revised. *ZAS Papers in Linguistics* 63, 1–36. doi: 10.21248/zaspil.63.2019.516
- Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: assessing performance of bilingual children. *Linguist. Educ.* 13, 175–197. doi: 10.1016/S0898-5898(01)00061-4
- Isbell, R., Sobol, J., Lindauer, L., and Lowrance, A. (2004). The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early Child. Educ. J.* 32, 157–163. doi: 10.1023/B:ECEJ.0000048967.94189.a3
- Kaderavek, J. N., and Sulzby, E. (2000). Narrative production by children with and without specific language impairment: oral narratives and emergent readings. *J. Speech Lang. Hear. Res.* 43, 34–49. doi: 10.1044/jslhr.4301.34
- Kail, M., and Hickmann, M. (1992). French children’s ability to introduce referents in narratives as a function of mutual knowledge. *First Lang.* 12, 73–94. doi: 10.1177/014272379201203405
- Kawar, K., Saiegh-Haddad, E., and Armon-Lotem, S. (2023). Text complexity and variety factors in narrative retelling and narrative comprehension among Arabic-speaking preschool children. *First Lang.* 43, 1–25. doi: 10.1177/01427237221149800
- Košutar, S., Kramarić, M., and Hržica, G. (2022). The relationship between narrative microstructure and macrostructure: differences between six- and eight-year-olds. *Psychol. Lang. Commun.* 26, 126–153. doi: 10.2478/plc-2022-0007
- Kunnari, S., and Välimaa, T. (2020). “Narrative comprehension in simultaneously bilingual Finnish-Swedish and monolingual Finnish children,” in *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives*, eds U. Bohnacker, and N. Gagarina (Amsterdam: John Benjamins), 150–170. doi: 10.1075/sibil.61.05kun
- Kunnari, S., Välimaa, T., and Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish-Swedish children. *Appl. Psycholinguist.* 37, 123–144. doi: 10.1017/S0142716415000442
- Kuvač Kraljević, J., Hržica, G., and Vdović Gorup, I. (2020). A comparative macrostructural analysis of narrative discourse in children with typical language development and children with developmental language disorder. *Drustvena Istrazivanja* 29, 453–470. doi: 10.5559/di.29.3.06
- Lever, R., and Sénéchal, M. (2011). Discussing stories: on how a dialogic reading intervention improves kindergartners’ oral narrative construction. *J. Exp. Child Psychol.* 108, 1–24. doi: 10.1016/j.jecp.2010.07.002
- Liles, B. Z. (1987). Episode organization and cohesive conjunctives in narratives of children with and without language disorders. *J. Speech Hear. Res.* 30, 185–196. doi: 10.1044/jslr.3002.185
- Liles, B. Z. (1993). Narrative discourse in children with language disorder and children with normal language: a critical review of the literature. *J. Speech Hear. Res.* 36, 868–882. doi: 10.1044/jslr.3605.868
- Lindgren, J. (2018). *Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish Children Aged 4–6*. Uppsala: Acta Universitatis Upsaliensis.
- Lindgren, J. (2022). The development of narrative skills in monolingual Swedish-speaking children aged 4 to 9: a longitudinal study. *J. Child Lang.* 49, 1281–1294. doi: 10.1017/S030500092100057X
- Lindgren, J., and Bohnacker, U. (2020). “Inferential comprehension, age and language: how German-Swedish bilingual preschoolers understand picture-based stories,” in *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives*, eds U. Bohnacker, and N. Gagarina (Amsterdam: John Benjamins), 61–98. doi: 10.1075/sibil.61.03lin
- Lindgren, J., and Bohnacker, U. (2022). How do age, language, narrative task, language proficiency and exposure affect narrative macrostructure in German-Swedish bilingual children aged 4 to 6? *Linguist. Approaches Biling.* 12, 479–508. doi: 10.1075/lab.20020.lin
- Lindgren, J., Tselekidou, F., and Gagarina, N. (2023). Acquisition of narrative macrostructure: a comprehensive overview of results from the multilingual assessment instrument for narratives. *ZAS Papers in Linguistics* 65, 111–132. doi: 10.21248/zaspil.65.2023.623
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*, 3rd Edn. Lawrence Erlbaum Associates.
- Mandler, J. M., and Johnson, N. S. (1977). Remembrance of things parsed: story structure and recall. *Cogn. Psychol.* 9, 111–151. doi: 10.1016/0010-0285(77)90006-8
- Maviş, I., Tunçer, M., and Gagarina, N. (2016). Macrostructure components in narrations of Turkish-German bilingual children. *Appl. Psycholinguist.* 37, 69–89. doi: 10.1017/S0142716415000429
- Merritt, D. D., and Liles, B. Z. (1987). Story grammar ability in children with and without language disorder: story generation, story retelling, and story comprehension. *J. Speech Lang. Hear. Res.* 30, 539–552. doi: 10.1044/jslr.3004.539
- Merritt, D. D., and Liles, B. Z. (1989). Narrative analysis: clinical applications of story generation and story retelling. *J. Speech Hear. Disord.* 54, 438–447. doi: 10.1044/jshd.5403.438
- Otwinowska, A., Mieszowska, K., Bialecka-Pikul, M., Opacki, M., and Haman, E. (2020). Retelling a model story improves the narratives of Polish-English bilingual children. *Int. J. Biling. Educ. Biling.* 23, 1083–1107. doi: 10.1080/13670050.2018.1434124
- Öztekin, B. (2019). *Typical and Atypical Language Development in Bilingual Turkish-Swedish Preschool Children*. Uppsala: Acta Universitatis Upsaliensis.
- Pavlenko, A. (2008). “Narrative analysis,” in *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, eds L. Wei, and M. G. Moyer (Malden, MA: Blackwell), 311–325. doi: 10.1002/9781444301120.ch18
- Peristeri, E., Andreou, M., Tsimpli, I. M., and Durrleman, S. (2020). “Bilingualism effects in the narrative comprehension of children with developmental language disorder and L2-Greek: links with language, executive function and theory of mind,” in *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives*, eds U. Bohnacker and N. Gagarina (Amsterdam: John Benjamins), 298–330.
- Reuterskiöld Wagner, C., Sahlén, B., and Nettelblad, U. (1999). What’s the story? narration and comprehension in Swedish preschool children with language impairment. *Child Lang. Teach. Ther.* 15, 83–93. doi: 10.1191/02656599971432141
- Ripich, D. N., and Griffith, P. L. (1988). Narrative abilities of children with learning disabilities and nondisabled children: story structure, cohesion, and propositions. *J. Learn. Disabil.* 21, 165–173. doi: 10.1177/002221948802100309
- Roch, M., Florit, E., and Levorato, C. (2016). Narrative competence of Italian-English bilingual children between 5 and 7 years. *Appl. Psycholinguist.* 37, 49–67. doi: 10.1017/S0142716415000417
- Schneider, P., and Dubé, R. V. (2005). Story presentation effects on children’s retell content. *Am. J. Speech Lang. Pathol.* 14, 52–60. doi: 10.1044/1058-0360(2005)007
- Schneider, P., Dubé, R. V., and Hayward, D. (2005). *The Edmonton Narrative Norms Instrument*. University of Alberta Faculty of Rehabilitation Medicine Website. Available online at: <http://www.rehabresearch.ualberta.ca/enni>
- Sheng, L., Shi, H., Wang, D., Hao, Y., and Zheng, L. (2020). Narrative production in mandarin-speaking children: effects of language ability and elicitation method. *J. Speech Lang. Hear. Res.* 63, 774–792. doi: 10.1044/2019_JSLHR-19-00087

- Stein, N. L., and Glenn, C. G. (1979). "An analysis of story comprehension in elementary school children," in *Discourse Processing: Multidisciplinary Perspectives*, ed R. Freedle (Norwood, NJ: Ablex), 53–120.
- Stein, N. L., and Policastro, M. (1984). "The concept of a story: a comparison between children's and teacher's viewpoints," in *Learning and Comprehension of Text*, eds H. Mandl, N. L. Stein, and T. Trabasso (Hillsdale, NJ: Lawrence Erlbaum), 113–155.
- Tribushinina, E., Irmawati, M., and Mak, P. (2022). Macrostructure in the narratives of Indonesian-Dutch bilinguals: relation to age and exposure. *Linguist. Approach. Biling.* 12, 540–570. doi: 10.1075/lab.20015.tri
- Walker, E. A., Harrison, M., Baumann, R., Moeller, M. P., Sorensen, E., Oleson, J. J., et al. (2023). Story generation and narrative retells in children who are hard of hearing and hearing children. *J. Speech Lang. Hear. Res.* 66, 3550–3573. doi: 10.1044/2023_JSLHR-23-00084
- Wehmeier, C. M. (2019). Development of narrative macrostructure in monolingual preschoolers in Germany and impact of socio-economic status and home literacy environment. *ZAS Papers in Linguistics* 62, 52–75. doi: 10.21248/zaspil.62.2019.443
- Wehmeier, C. M. (2020). "Narrative comprehension and its associations with gender and nonverbal cognitive skills in monolingual and bilingual German preschoolers," in *Developing Narrative Comprehension: Multilingual Assessment Instrument for Narratives*, eds U. Bohnacker, and N. Gagarina (Amsterdam: John Benjamins), 270–296. doi: 10.1075/sibil.61.09weh
- Westby, C. E. (2012). "Assessing and remediating text comprehension problems," in *Language and Reading Disabilities, 3rd edn*, eds A. G. Kamhi, and H. W. Catts (Boston, MA: Pearson), 163–225.
- Yang, W., Chan, A., and Gagarina, N. (2023). Left-behind experience and language proficiency predict narrative abilities in the home language of Kam-speaking minority children in China. *Front. Psychol.* 13:1059895. doi: 10.3389/fpsyg.2022.1059895

APPENDIX

TABLE A1 Studies using MAIN to investigate the effect of task type (telling vs. retelling).

Study	Groups	N	Ages	SC	SS	Task order	Retell pres	Results
Kunnari and Välimaa (2020)	Mo Finnish Bi Swedish-Finnish	16 + 16	5–6	X		Tell-Retell	EXP	ns
Kunnari et al. (2016)	Mo Finnish Bi Swedish-Finnish	16 + 16	5–6		X	Tell-Retell	EXP	Retell > Tell only for Bi in Finnish
Kuvač Kraljević et al. (2020)	Mono Croatian TD + DLD	20 + 20	mean 6;6	X		Tell-Retell	HP	Retell > Tell
Maviş et al. (2016)	Bi Turkish-German	13	5–7	X	X	Tell-Retell	EXP	ns
Otwinowska et al. (2020)	Mono Polish Bi Polish-English	75 + 75	3–7	X	X	Tell-Retell	EXP	Retell > Tell
Roch et al. (2016)	Bi Italian-English	62	5–7	X	X	CB	HP	Retell > Tell
Sheng et al. (2020)	Mo Mandarin Chinese TD + “at risk”	21 + 21	mean 5;8	X		Retell-Tell	EXP	ns for TD, Retell > Tell for “at risk”
Wehmeier (2019)	Mono German	198	4.5–6	X	X	Retell-Tell	EXP	Retell > Tell
Wehmeier (2020)	Mono German Bi German	199 + 66	4.5–6	X		Retell-Tell	EXP	Retell > Tell

All studies used Cat/Dog in retelling and Baby Birds/Baby Goats in telling, except Wehmeier (2019, 2020) who used Cat in retelling and Baby Birds in telling; SC = story comprehension; SS = story structure; Retell pres = mode of presentation of the model story in the retelling task (EXP = read by the experimenter; HP = prerecorded, presented over headphones); Mo = monolingual; Bi = bilingual; CB = order of the tasks was counterbalanced; ns = non-significant.

TABLE A2 Macrostructural components in the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015).

Component	Description
Setting	Time and place of the events
Internal State as Initiating Event (IS as IE)	What does the character perceive/feel that sets the story events in motion?
Goal (G)	What does the character want?
Attempt (A)	What does the character do (in order to reach the goal)?
Outcome (O)	What is the result? what happens?
Internal State as Reaction (IS as R)	What does the character feel (in response to the outcome)?

TABLE A3 Overview of the ten comprehension questions, MAIN Cat.

Question number	Question
D1. episode 1 goal	Why does the cat jump/leap forward?
D2. episode 1 IST	How does the cat feel?
D3. episode 1 IST rationale	Why does the cat feel [answer D2]?
D4. episode 2 goal	Why does the boy hold the fishing rod in the water?
D5. episode 2 IST	How does the boy feel?
D6. episode 2 IST rationale	Why does the boy feel [answer D5]?
D7. episode 3 goal	Why does the cat grab the fish?
D8. theory of mind IST	Imagine that the boy sees the cat. How does the boy feel?
D9. theory of mind IST rationale	Why would the boy feel [answer D8]?
D10. overall plotline question	Will the boy be friends with the cat? Why?

IST = internal state term.