Developing Narrative Comprehension

Multilingual Assessment Instrument for Narratives

EDITED BY
Ute Bohnacker
Natalia Gagarina

STUDIES IN BILINGUALISM

61

JOHN BENJAMINS
PUBLISHING COMPANY

Developing Narrative Comprehension

Studies in Bilingualism (SiBil)

ISSN 0928-1533

The focus of this series is on psycholinguistic and sociolinguistic aspects of bilingualism. This entails topics such as childhood bilingualism, psychological models of bilingual language users, language contact and bilingualism, maintenance and shift of minority languages, and socio-political aspects of bilingualism.

For an overview of all books published in this series, please see
benjamins.com/catalog/sibil

Volume 61

Developing Narrative Comprehension
Multilingual Assessment Instrument for Narratives
Edited by Ute Bohnacker and Natalia Gagarina

Developing Narrative Comprehension

Multilingual Assessment Instrument for Narratives

Edited by

Ute Bohnacker
Uppsala University

Natalia Gagarina
Leibniz-ZAS and Uppsala University

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/sibil.61

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2020040620 (PRINT) / 2020040621 (E-BOOK)

ISBN 978 90 272 0808 8 (HB)
ISBN 978 90 272 6034 5 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Acknowledging our reviewers

Mehmet Akinçi
Ayhan Aksu Koç
Carmit Altman
Elma Blom
Tessel Boerma
Christiane Bongaertz
Nicola Botting
Cécile de Cat
Vicky Chondrogianni
Sveta Fichman
Ewa Haman
Kristina Hansson
Svetlana Kapalková
Nihan Ketrez
Sari Kunnari
Aylin Küntay
Carolyn Letts
Chiara Levorato
Josefin Lindgren
Leena Mäkinen
Karolina Mieszkowska
Julie Lynch
Marilyn Nippold
Philippe Prévost
Christina Reuterskiöld
Maja Roch
Phyllis Schneider
Nathalie Topaj
Teresa Ukrainetz
Taina Välimaa
Monique Visser
Marleen Westerveld

Cross-linguistic development of narrative comprehension from A to Z

Ute Bohnacker[1] and Natalia Gagarina[1,2]
[1]Uppsala University / [2]Leibniz-Centre General Linguistics (ZAS)

1. Introduction

Imagine a series of pictures, in which, close to some watery expanse, a cat is falling into a bush, a butterfly is flying away after the cat has unsuccessfully tried to catch it, and a boy is passing by with a ball, a fishing rod and a bucket of fish. Startled and surprised by the cat struggling in the bush, the boy lets go of his ball, which rolls down into the water. The boy then goes down on his knees and moves his fishing rod towards the ball. Why does the boy do this? The answer seems to be clear and simple – in order to get his ball back.

Imagine further that you also see that the cat in the following pictures, unbeknownst to the boy, takes and eats the fish in the bucket that the boy left on the shore. How will the boy feel if/when he sees that his fish are being eaten by the cat? Putting yourself in the boy's shoes, an immediate answer that comes to mind is that the boy would feel *bad, sad, angry* or *mad*. Answering that the boy would feel *fine, good, happy, satisfied* or *pleased* seems to be out of place, illogical and incorrect. However, if you are a generous person and think a bit more about the whole situation, you might in fact feel sorry for the cat, not begrudge the hungry cat the fish and forgive the cat, in which case an answer such as *fine* or *good* would also be appropriate.

Both of the above questions tap into the visual content of the picture series and into cause-effect relationships involving story protagonists, yet they differ in their degree of complexity as far as inferencing and understanding of the situation is concerned. So how do children understand different types of pictorial story content at different ages? What causes difficulties, and what is easier to comprehend? And what does this tell us about children's comprehension of the story as a whole? This volume offers some answers to these questions and presents narrative comprehension results from 812 children in different settings around the world, from age 2;10 to age 9;9, the majority being between 4 and 7 years, for ten different languages and

https://doi.org/10.1075/sibil.61.01boh

more than a dozen different language pairings. Throughout, the same experimental methodology, the *Multilingual Assessment Instrument for Narratives* (MAIN, Gagarina et al., 2012, 2019) is used.

The main goal of the present volume is to take several steps forward in our understanding of children's narrative comprehension skills. MAIN provides a tool for narrative comprehension that can be used with children already from around age 3. For the analysis of narrative production (storytelling and/or retelling), a number of materials and methods have been available for some time, including the well-known *Frog story*, which has been comparatively studied in numerous languages around the world (Berman & Slobin, 1994). For the analysis of narrative comprehension, a tool has been lacking. MAIN includes such a tool, and not only can it be used in different languages and with both monolinguals and bilinguals, but it also brings the age down. Children's narrative abilities can thus be investigated at an earlier age and at lower proficiency levels than what is generally possible with narrative production tasks. MAIN story comprehension (as reflected in children's responses to comprehension questions) can be used even when children are not yet able to produce a comprehensible fictional narrative themselves – this holds for instance for many typically developing children aged 3 to 4 years, and for child second language learners of any age with little exposure and low-level language skills, as well as for children with developmental language disorders. Of course, the tool can also be used with more proficient and older children. And as MAIN contains near-identical comprehension tasks for structurally parallel stories, narrative comprehension can straightforwardly be compared in the two languages of bilingual children.

MAIN allows us to quantify children's story comprehension abilities, at least to some extent, and these comprehension scores can then be compared. One of the goals of this volume is to reach a better understanding of the children's *developmental trajectory* for story comprehension. Since MAIN is not a standardised test, the volume cannot provide age norms, but it can provide reference data and *a sense of what can be expected* at which age for story comprehension, similarly to what the cross-linguistic Frog story project of Berman, Slobin and colleagues has done for narrative production around the world.

The contributions in this book use the same methodology across languages and participant groups, and the same set of theoretically motivated comprehension questions. Taken together, they aim to provide a sense of which story aspects can be expected to be understood by children at which age and under which circumstances. As both monolingual and bilingual children and both languages of bilinguals are investigated, the volume also sheds light on other questions, for example, how does story comprehension 'behave' in two languages, i.e., is it similar or not?

The contributions also illustrate the variety of analyses that MAIN allows us to do: Which aspects of a story are comprehended more easily (or earlier) by children than others, and why? How does modality (e.g. first listening to a story and retelling it vis-à-vis generating a story (telling) without hearing it first) affect story comprehension? How is story comprehension affected by language skills, such as receptive and/or expressive vocabulary knowledge? How is story comprehension influenced by non-linguistic factors, such as chronological age and a range of environmental factors, such as age of onset, language exposure, socio-economic status, exposure to narratives or literacy expectations? How is story comprehension related to general cognitive skills? And how is story comprehension related to story production, since story production has long been the default, standard way of investigating narrative abilities in children?

Last but not least, the book aims to make a methodological contribution. Since MAIN is a relatively newly developed and revised tool (Gagarina et al., 2012, 2019), and few publications on narrative comprehension with MAIN exist, an additional goal is to put our tool to the test. How easy is it to administer, elicit, score, and compare results across groups and languages? Are there unintended task or story effects, or item effects? Most of all, how well does MAIN work to capture and assess children's narrative comprehension abilities? Are there floor effects, increases, plateaus, and/or ceiling effects?

In the remainder of this introductory book chapter, we discuss some of the theoretical and empirical issues connected with the comprehension of picture stories, both visually presented stories, and visual stimuli accompanied by oral stimuli (i.e. listening to a story). We define what narrative comprehension is and review some of the theoretical background on narrative comprehension. We present our model of inference-based narrative comprehension and describe the elicitation tool used in this book. We then trace the developmental trajectory of narrative comprehension, as it manifests in the results of the chapters in this volume. Finally, we specify how MAIN as a tool for story comprehension can help us advance our knowledge of language acquisition in general and of language comprehension in particular.

2. Narrative comprehension and how it can be understood

2.1 Theoretical background

All attempts to investigate narrative comprehension start with the assumption that events in a story are interconnected in some fashion (Thompson & Myers, 1985, p. 1143).

But what is narrative comprehension exactly? Comprehension of any type of discourse means that the comprehender constructs a coherent representation of what the discourse is about in his or her mind (e.g. Mandler & Johnson, 1977;

Kintsch & van Dijk, 1978; Johnson & Mandler, 1980; Kintsch, 1988; Stein & Glenn, 1979; Trabasso, Secco, & van den Broek, 1984; Graesser, Singer, & Trabasso, 1994; Hayward, Schneider, & Gillam, 2009). Narratives or stories are a particular type of discourse, and one of their fundamental characteristics is that they have a causal event structure. Comprehenders of a narrative or story construct a coherent and meaningful mental representation of what the story is about, and in order to do so need to identify and infer meaningful relations, including the motivational and causal relations between events. According to a number of researchers, this representation takes the form of a mental network that mirrors the causal (and other) relations between events that the comprehender has recognised or inferred (e.g. Kintsch & van Dijk, 1978; Trabasso, Secco, & van den Broek, 1984; Trabasso & Nickels, 1992).

Comprehenders of stories construct their mental representations on the basis of two things: firstly, the information provided in the story, and secondly, the comprehender's background knowledge and expectations arising from this knowledge. Background knowledge includes knowledge of facts of the world, how social interaction works, cause-effect relationships, and knowledge of narrative schemata.

Information provided in the story can be presented in various ways via different media. Purely verbally presented stories can be narratives told or read aloud to the listening comprehender, or written texts that the comprehender reads. Audio-visual presentations of stories are, for instance, films or oral stories told accompanied by pictures. Other stories are presented purely visually, such as silent films, nonverbal video clips, pictures, murals or frescoes, picture sequences or wordless picture books. Importantly, neither visually presented nor verbally presented stories make all aspects of a story explicit. Story comprehension thus always necessitates inferencing, for example of the goals and emotions of story characters. To draw inferences means to integrate various types of information, which may be taken from different sources, into a conclusion (e.g. Graesser et al., 1994, pp. 373–376; Levinson, 1983, pp. 21–22).

On this view, narrative comprehension is thus not the ability to recall as many story events or facts as possible. Rather, narrative comprehension is the identification of causal, hierarchical and thematic relations between events and clusters of events, and the encoding of these relations in memory, resulting in an understanding of the overall plot or point of the story.

Whilst even very young children are able to identify and infer meaningful relations between events to some degree and create mental network representations of these events (Thompson & Myers, 1985; Trabasso & Nickels, 1992; van den Broek, Kendeou, Kremer, Lynch, Butler, & Lorch, 2005), their narrative comprehension is generally not as developed as that of older children and adults. This is hardly surprising, as children with every year of their lives accumulate new experiences and knowledge of facts about the world, as well as social knowledge of how humans interact. With increasing experience and background knowledge, they can then make more efficient use of their attentional and working memory capacities in story comprehension. Their inferential processes do not only connect concrete, physical events that occur closely together, but increasingly also connect abstract, internal events and/or events that are more distant (van den Broek et al., 2005, pp. 116–118). Cognitive development and cognitive maturity are not only reflected in the way children tell and retell stories, but also in the types of answers children provide in response to story comprehension questions.

Narrative comprehension is a prerequisite for narrative production: If children do not understand cause-effect relationships, plotlines, intentions, thoughts and feelings of story characters, they will not be able to convey this information to the listener during storytelling either (Stein & Glenn, 1979; Shapiro & Hudson, 1991, p. 115; Trabasso & Nickels, 1992; Astington & Pelletier, 2005, p. 327; Burris & Brown, 2014). In the world of stories, autonomous agents (i.e. story characters) have intentions and goals, and act and react in relation to these goals. Fictional stories take place in an imagined world, but the agents, their intentions and goals, their actions and reactions are shaped by our own experiences. To understand the story events depicted or described, comprehenders often assume the story character's perspective, mentally simulating and representing their internal states (Mar, 2004, p. 1416).

In story comprehension and storytelling, children must thus make inferences about the internal (or mental) states of others (e.g. Bishop, 1997; Letts, & Leinonen, 2001; van den Broek et al., 2005). This requires theory of mind, that is the awareness of mental states and "the ability to use this awareness in interpreting, explaining and predicting the behavior of one self and others" (Astington & Pelletier, 2005, p. 313). Theory of mind plays an important role in children's socio-cognitive development, as it concerns the understanding of (real) persons, as well as of fictional characters, as psychological beings. Here it has been found that even though young children rarely mention internal states in their storytelling, they are able to answer inferential questions that directly probe these internal states, e.g. the goals and emotions of story characters (Stein & Glenn, 1979; Trabasso, Stein, Rodkin, Munger, & Baughn, 1992; Wenner, 2004; Lynch & van den Broek, 2007; Hayward, Schneider, & Gillam, 2009; Tompkins, Guo, & Justice, 2013; Bohnacker, 2016).

Narrative comprehension involves interrelated cognitive skills that allow a child to build a coherent representation of a story. As these skills also play an important role in reading and understanding text, narrative comprehension is often seen as a prerequisite for reading achievements at school age (Paris & Paris, 2003; van den Broek et al., 2005; Lynch, van den Broek, Kremer, Kendeou, White, & Lorch, 2008).

2.2 Comprehension of stories in different modalities

As mentioned before, comprehenders encounter stories in different modalities. The literature on narrative comprehension is dominated by studies of reading comprehension of written texts or of listening comprehension of aurally presented texts (e.g. Thompson & Myers, 1985; van den Broek, 1997; Paris & Stahl, 2005; Lynch & van den Broek, 2007; Kendeou, Bohn-Gettler, White, & van den Broek, 2008; Florit, Roch, & Levorato, 2011; Filiatrault-Veilleux, Bouchard, Trudeau, & Desmarais, 2015). Both reading and listening comprehension rely on verbally presented stories. MAIN is different in this regard, since it is centred on the visual presentation of the story (see below).

In a number of studies and materials, for instance the well-known Bus Story Test (Renfrew, 1969), the child first listens to a story, with or without pictorial support, and then has to retell or recall the story. Recall is taken as an indirect measure of story comprehension. However, as the child has to hold large chunks of aurally presented information in working memory, this type of story comprehension assessment in effect becomes a memory task (Boudreau, 2007; Dodwell & Bavin, 2008; Bohnacker & Lindgren, in press).

A similar conflation occurs when the child first listens to the story and is then asked comprehension questions (for instance, the well-known Tiger's whisker story (Stein & Glenn, 1979; Trabasso et al., 1984; Gutiérrez-Clellen, 2002), and many of the materials in a recent scoping review by Filiatrault-Veilleux et al., 2015). To score correct on such questions, the child must recall details that were previously mentioned, or, if the question is an inferential one, draw inferences. Yet if the to-be-recalled facts are only presented aurally, story comprehension processes are again heavily weighed down by memory demands.

In the literature on children's narrative comprehension, studies that employ a purely visual presentation mode for stories are relatively uncommon, but they do exist (e.g. Ellis Weismer, 1985; Bishop & Adams, 1992; Paris & Paris, 2003; Hayward et al., 2009). Here, the child does not hear the story first but studies a picture sequence and is then asked comprehension questions that assess inferencing abilities. Surprisingly, in virtually all existing studies, the pictorial stimuli are removed before the comprehension questions are asked so that the child is no longer able to look at and refer to the pictures (e.g. Bishop & Adams, 1992; Letts & Leinonen, 2001; Paris & Paris, 2003; van den Broek et al., 2005; Lynch et al., 2008; and the studies in the scoping review by Filiatrault-Veilleux et al., 2015). Such pictorial story comprehension assessment in effect becomes a memory task. If, by contrast, the stimulus pictures are kept in full view of the child so that they can be referred to during comprehension questions, attentional resources and working memory are taxed much less. This alternative approach is rarely used though; a notable exception is

Hayward et al. (2009). The above discussion underscores the need to be sensitive to the task demands involved in assessing children's narrative comprehension abilities.

In MAIN, stories are never presented only in the listening mode, but always via picture sequences. Importantly, these pictorial stimuli stay in full view of the child throughout the procedure. MAIN includes both storytelling and story comprehension tasks with picture sequences as a base (see Section 4). Story comprehension is assessed via probe questions that require the child to make inferences from the pictures and to verbalise this understanding. Story comprehension in MAIN can be assessed in three ways: (1) via Model story, (2) via Retelling, and (3) via Telling. All three have been used by the various contributors to the present volume. In (1), Model story, the child first studies the MAIN picture sequence, then listens to the story script with the pictures remaining in full view, and finally answers comprehension questions, with the pictures still in full view. In alternative (2), Retelling, the child first studies the picture sequence, then listens to the story script with the pictures remaining in full view, retells the story with the pictures still visible, and finally answers the comprehension questions, again with the pictures in full view. In alternative (3), Telling, the child first studies the picture sequence, does not listen to the story, but rather generates and tells his/her own story from the pictures, which stay visible to the child during this process. Then, the comprehension questions are asked, with the pictures still in full view of the child. The procedure of keeping the MAIN pictorial stimuli visible throughout has been designed to minimise the conflation of story comprehension with the assessment of memory skills.

All comprehension questions asked in MAIN are inferential; no 'factual' (or 'literal') questions are asked. Some previous studies have investigated children's comprehension of narratives with different types of narrative comprehension questions (e.g. Bishop & Adams, 1992; Letts & Leinonen, 2001; Gutiérrez-Clellen, 2002; Paris & Paris, 2003); often a mix of factual and inferential questions is used. Factual narrative comprehension questions target recall (memory) of factual details that have either been presented verbally (listening comprehension) or visually (stimulus pictures/film). Examples of factual questions are: *Which colour were the lady's shoes? – Red; Where did they live? – In the middle of the forest; How long did they stay on the island? – 15 months*. Inferential questions target comprehension of story aspects that were not explicitly mentioned or depicted and must be induced/inferred. Examples of inferential questions are: *How does the boy feel?, Why do you think the boy felt bad?, Why does the bird bite the fox's tail?* It is such inferential comprehension questions that MAIN focuses on (see Section 4.3).

Inferential story comprehension tasks usually necessitate that the child during processing takes into account several pictures in the sequence, and not just one single picture. Content is compositional and connected. Thus the same character may appear again and again throughout the sequence, and events depicted earlier on in the sequence often lead to other events and reactions later on. Being aware of and recognising these connections between pictures is a prerequisite for making the correct inference. So as not to make this unnecessarily hard for the child, the picture sequence is kept in full view of the child when the MAIN inferential comprehension questions are asked.

3. Our model of inference-based visual narrative comprehension

Story comprehension is the construction of a coherent mental representation of what the story is about, as discussed in Section 2. Since different types of discourse have their own organisation (schemata), story comprehension involves constructing or activating a narrative schema that can provide a coherent representation of story events and their protagonists. Stories are made up of episodes with a number of different story components. However, in our view, narrative comprehension is not the ability to recall as many story events or facts as possible, nor the ability to describe as many story events or details depicted in the visual stimulus materials. Rather, comprehenders identify, recognise and infer meaningful relations, particularly causal relations between events. For this, they need to assume the story characters' perspectives, by mentally simulating their intentional and emotional states, at least momentarily. Comprehenders thus make mental representations of the story characters' internal states.

In our view then, a good theory or model of narrative comprehension, as well as its practical application, should focus on the story character's internal states and on the motivational and emotional cause-effect relationships between events that must be inferred. Here, the impact of other factors, such as extraneous memory load, on inferential comprehension should be kept to a minimum.

Our proposed focus on inferencing means that in narrative comprehension little emphasis, if any, needs to be put on remembering literal or factual aspects of the story. Story components or points of detail that are explicitly mentioned in the verbal version or are shown in the pictorial stimuli thus need not be queried (e.g. with *what/where/when/how big/how long*/etc. questions).

Instead, questions of the type *Why does X do Y?* target motivations (goals) and cause-effect relationships, whilst questions of the type *How (do you think) X feels?* target internal states of characters. Follow-up questions of the type *Why do you think/say that …?* can shed further light on the inferencing process, as they query the reasons or rationale behind the child's answers to *why* questions and *how … feel* questions. Comprehension of the story as a whole can be probed in different ways, where follow-up *why*-questions (*Why do you think/say that …?*) target the reasons or rationale behind a certain answer, thus elucidating the inferencing process.

Working memory certainly plays a role in inferencing processes, but in our view the memory load should not be unnecessarily increased in narrative comprehension tasks, so as not to conflate comprehension with recall. We therefore advocate the use of carefully designed visual stimulus materials in narrative comprehension tasks with children, which should remain in full view of the child throughout to reduce task demands, as discussed in Section 2.

This model of narrative comprehension is put into concrete practice with the comprehension task of MAIN (see Section 4.3).

4. The MAIN materials

4.1 Background: The 'birth' of MAIN

MAIN contains four structurally parallel fictional picture sequences and a standardised protocol to assess both narrative production and comprehension in children aged 3–10 years. Since there are four picture sequences, bilingual children can be tested in both their languages with comparable stimuli, without reusing the same story. This reduces training effects and boredom, which otherwise easily occur when the same material is used with the same child several times.

MAIN was first published in 2012/2013. Several years of theory development and material construction preceded its launch. As part of the European COST Action IS0804, the Working Group on Narrative and Discourse first evaluated existing models of story organisation (e.g. Labov & Waletsky, 1967; Labov, 1972; Mandler & Johnson, 1977; Johnson & Mandler, 1980; Stein & Glenn, 1979; Peterson & McCabe, 1983; Stein & Policastro, 1984; Trabasso & Nickels, 1992; Berman & Slobin, 1994; Westby, 2005). The core group of MAIN developers then created a multi-dimensional model of story organisation, where goals are central to narrative structure. This model of story organisation was translated into child-appropriate real-life plots, which are also cross-culturally and cross-linguistically robust. The episodes in these plots are all clearly goal-based. The plots were visualised as picture stories and piloted for 15 languages.[1] This entire process was accompanied by the

1. MAIN is part of the LITMUS (Language Impairment Testing in Multilingual Settings) community. LITMUS includes a battery of tests that were developed in connection with COST Action IS0804 *Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment* (2009–2013), funded by the European Cooperation in Science and Technology (COST). Since then, the MAIN network has expanded beyond the LITMUS community of researchers and practitioners. Interdisciplinary teams of theoretical linguists, psycholinguists, psychologists, clinicians and educators are engaged in the MAIN network across countries and continents. Some of them have contributed to this volume.

development of elicitation procedures for story production and story comprehension, assessment protocols and scoring guidelines. All this work led to the birth of MAIN, as documented in the manual (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012), published in *ZAS Papers in Linguistics* (*ZASPiL* 56), together with 26 language versions of MAIN.

Three years later, in 2015, first results for 17 languages and 14 different language pairs (Gagarina et al., 2015) were published in the volume *Assessing multilingual children* (Armon-Lotem, de Jong, & Meir (Eds.), 2015). In 2016, a special issue of *Applied Psycholinguistics* "Narrative abilities in bilingual children" with 7 original research articles presented new results for MAIN. A number of publications followed.

In 2019, the Revised version of MAIN in English (as a base for all language adaptations) (Gagarina, Klop, Kunnari, Tantele, Välimaa, Bohnacker, & Walters, 2019) was published in *ZAS Papers in Linguistics* (*ZASPiL* 63), together with revised versions in German, Russian, Swedish, as well as Turkish for the bilingual Turkish-Swedish population in Sweden. These revisions were the result of intensive collaboration between Ute Bohnacker's research group in Uppsala and Research Area 2 at the Leibniz-ZAS in Berlin, led by Natalia Gagarina. The two groups worked through 2,500 transcribed oral narrative texts and more than 24,000 responses to MAIN comprehension questions of monolinguals and bilinguals in Sweden, Germany and Russia, in order to improve the guidelines, elicitation and scoring procedures.

4.2 The picture stories

MAIN provides four stimuli sets of coloured picture sequences of six pictures each to elicit four stories with identical overall story and episode structure: *Cat, Dog, Baby Birds* and *Baby Goats*. In contrast to conventional picture books (and also in contrast to *Frog where are you?* (Mayer, 1969) and related Frog stories), the MAIN stimulus materials were carefully constructed to be of comparable complexity, with story grammar as a theoretical base (for details see Gagarina et al., 2012, pp. 21–48).

Each of the four MAIN stories is made up of three episodes. Episodes are a chronologically ordered group of events within a larger narrative that are conceptually connected to a specific goal of a story character (Stein & Glenn, 1979). Each episode consists of an internal state (IST) as initiating event, goal, attempt, outcome and internal state as reaction of the characters. An 'initiating event' is an action or idea in the beginning of each episode. The protagonist sets a 'goal' in response to the initiating event and the effort to achieve this goal is called 'attempt'. The 'outcome' is the result of the attempt or action, which can be either success or failure. An 'internal reaction' is the (emotional) response of the character to the outcome.

Two of the four picture sequences, the *Cat* and *Dog* stories, are parallel in plotline, story grammar and length. Both Cat and Dog have three main characters; the only difference is the characters and objects in the stories. The Cat and Dog story picture sequences are shown in Figure 1.

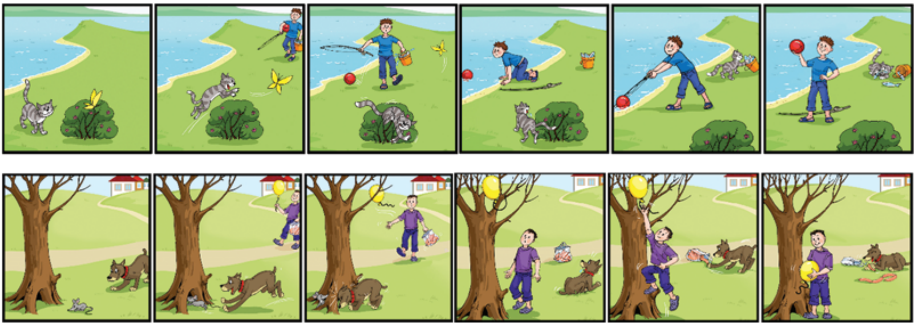

Figure 1. Small-scale copy of the MAIN Cat picture sequence (top) and the Dog picture sequence (bottom) (Gagarina et al., 2012, 2019). The original pictures are in colour and size 9×9 cm (reproduced with permission from the publisher)

The Cat story consists of three partially overlapping episodes with three characters, a cat, a butterfly and a boy, and some objects, which include a bush, a ball and a bucket of fish. In the first episode, a cat wants to catch a butterfly and as an attempt, the cat jumps on the butterfly but gets stuck in a bush. The second episode starts when a boy carrying a ball, a fishing rod and a bucket of fish comes and sees the cat and butterfly (as also mentioned at the beginning of this chapter). Surprised, the boy accidentally drops his ball, which rolls into the water, and he tries to get it out again with the fishing rod. In the third episode, the cat sees the fish in the bucket that the boy put down on the ground, and it decides to steal the fish. In the end, the cat takes the fish and eats them. In the meantime, the boy manages to take his ball and is happy about that. He has not yet realised that the cat has stolen his fish.

In the Dog story, the plotline is identical to the Cat story, only with different characters and objects. The characters are a boy, a mouse and a dog, and the objects include a tree, a balloon and a bag of sausages. In Episode 1, a dog sees a mouse and tries to catch it. The mouse gets in underneath a tree trunk, the dog jumps but misses the mouse and as a result hits its head on the tree. In Episode 2, a boy comes with a balloon on a string and a bag of sausages in his hand and sees the two animals. Surprised, he accidentally lets go of his balloon, which flies up into the tree. The boy puts his bag of sausages down on the grass and tries to reach his balloon by climbing the tree. Seeing the sausages, the dog decides to steal them in Episode 3. While the dog grabs and eats the sausages, the boy gets the balloon back and is happy about that. He has not yet noticed that the dog has stolen his sausages.

The other two picture sequences, the *Baby Birds* and *Baby Goats* stories, are also parallel in plotline, story grammar and length. Both consist of three episodes, and they have five animal characters each, which makes them somewhat more complex than Cat and Dog as far as the number of characters is concerned, with a slightly different episodic organisation. Baby Birds and Baby Goats are both about a family (of birds/goats), an attacker, and a rescuer. The Baby Birds and Baby Goats picture sequences are shown in Figure 2 below.

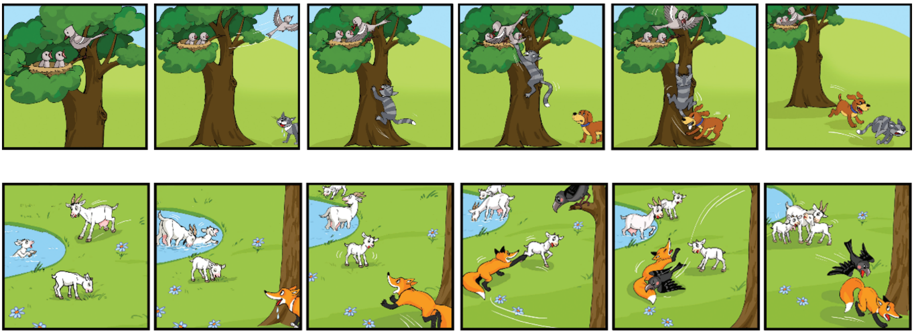

Figure 2. Small-scale copy of the MAIN Baby Birds picture sequence (top) and the Baby Goats picture sequence (bottom) (Gagarina et al., 2012, 2019). The original pictures are in colour and size 9×9 cm (reproduced with permission from the publisher)

In Baby Birds, the first episode begins with a nest of hungry baby birds up in a tree, crying for food to their mother/parent. The mother bird flies away to bring a worm for them to eat. Meanwhile, in Episode 2, a cat sees that the baby birds are alone and decides to catch them. The cat starts to climb the tree and grabs at them. In the third episode, a dog sees the cat and decides to rescue the baby birds, so the dog jumps and grabs the cat's tail, pulls it down and the cat runs away from the dog. The birds are together and safe.

In the Baby Goats story, although the general episodic structure and the number of characters are identical to Baby Birds, the content of the first episode is somewhat different. Here, a family of goats is in a meadow by a lake, one baby goat is drowning in the lake and the mother/parent goat runs down into the water to rescue it. Meanwhile in Episode 2, a fox sees the other baby goat feeding on the grass and wants to catch it. The fox jumps out from behind a tree towards the baby goat and grabs its hind leg. In Episode 3, a crow sees this situation and intervenes by attacking the fox and rescues the baby goat. The fox runs away from the crow. The baby goat is safe, and the goat family is reunited.

The stories of the picture sequences have been carefully designed to be parallel in macrostructure (overall organisation of narrative content, i.e. story structure and episodic structure) and length (for details see Gagarina et al., 2012, pp. 10–48).

4.3 The MAIN comprehension task

The present volume is about the narrative comprehension task included in MAIN. This comprehension task is picture-based and theoretically grounded, targeting the understanding of core components that are not directly depicted in the stories but must be inferred. Ten scripted *why* and *how* questions (D1–D10) for each picture series probe the children's inferencing abilities concerning the goals (intentions), thoughts and feelings of the story protagonists (e.g. *Why does the cat jump forward?, How does the cat feel?, Why do you think the cat is feeling …?*) and overall story comprehension. The answers to the 10 questions in the comprehension task are scored one point each, with a maximum of 10 points per story.

There is a standardised procedure for administering MAIN (for details, see Gagarina et al., 2012, 2019). The narrative comprehension task is carried out *directly after* the narrative production or listening task (i.e. after model story, retelling or telling), which is performed while looking at the pictorial stimuli. In the model story mode, the child first hears the story (typically *Cat* or *Dog*) from an adult experimenter to get a model of how a story is told while looking at the pictures, and then answers the comprehension questions. In the retelling mode, the child hears the story from an adult while looking at the pictures and then tells the same story while looking at the pictures. In the telling mode, the child generates the story (typically *Baby Birds* or *Baby Goats*) without any prior model of the story provided by the adult. For familiarisation, the entire picture sequence is first shown to the child and then folded up again. Then, the experimenter opens the pictures two by two and lets the child tell the story until the entire picture sequence is unfolded.

When the comprehension questions are asked, the pictures can be seen by both child and experimenter, for joint visual attention. Where appropriate, the experimenter points to the respective picture or story character that a question is being asked about. As the pictures remain in full view, the comprehension task does not probe the child's recall of events (memory). Rather, the questions require the child to make inferences from the pictures and state why s/he has come to a certain conclusion. The questions thus assess inferencing, i.e. how well the child is able to interpret physical and psychological cause-effect relationships and recognise characters' goals, the reasons for these goals, and the reactions following attempts to reach the goals, as well as the understanding of the story as a whole. The comprehension task is particularly informative in case of poor story telling performance, since children may know more about narrative structure than their own narratives can display.

Importantly, the questions used for the four MAIN stories are designed to be parallel, asking for the same types of essential story information. This enables comparisons across the MAIN stories. By contrast, many earlier studies of comprehension asked an ad hoc set of questions that were tied to particular stories but did not attempt to collect the same kind of information across stories. Since Cat and Dog are very similar, the comprehension questions for these stories are practically identical. Baby Birds and Baby Goats are also similar, as are the comprehension questions to these stories (for critical discussion, see Bohnacker, Öztekin, & Lindgren, this volume).

For all MAIN stories, three questions (D1, D4 and D7) probe understanding of the goals of the main characters in the three episodes (see Table 1). Six questions (D2, D3, D5, D6, D7 and D8) probe internal states that either initiate an event or are a reaction to an outcome of an event. Questions D2 and D5 query the internal states of the characters in Episode 1 and 2. Question D8 queries a character's internal state in Episode 3, but does so with a theory of mind question, testing the child's understanding by a 'what if' scenario that is not actually shown in the picture series. Questions D3, D6 and D9 are follow-up questions to the preceding questions about a character's feelings. These follow-up questions ask *why* a character feels a certain way. If the child does not answer the preceding question correctly, the follow-up question is not asked according to protocol. The tenth and final question, D10, tests whether the child can infer meaning about the story as a whole (overall plotline). It is made up of both a yes/no question and a *why* question, and only scored as correct if the child answers both questions correctly.

The ten questions and examples of correct and incorrect responses are illustrated for the Cat story in Table 1 below. As shown in the table, the questions are scripted, the experimenter is given supporting information about the type of question s/he is asking (e.g. IST as reaction) and is instructed when to guide the child's gaze, and where to, by pointing to the relevant character or picture (e.g. "point to picture 3").

Table 1. Cat story comprehension questions and examples of responses (MAIN-Revised, Gagarina et al., 2019)

	Question asked by experimenter	**Examples of correct responses (Scored 1 point)**	**Examples of wrong responses (Scored 0 points)**
0	Did you like the story?	Warm-up question, not scored	
D1.	**Why does the cat jump/leap forward?** *(point to pictures 1–2)* (Episode 1: Goal)	Wants to get/ catch/ chase the butterfly/ to play with the butterfly Wants the butterfly (In order) to + VERB (get, take) the butterfly	Is leaving/ running/ wanted to jump Cats are always jumpy/ running
D2.	**How does the cat feel?** *(point to picture 3)* (IST as reaction)	Angry/ bad/ disappointed/ hurt/ in pain/ not good/ not comfortable	Good/ happy

Table 1. (*continued*)

	Question asked by experimenter	Examples of correct responses (Scored 1 point)	Examples of wrong responses (Scored 0 points)
D3.	*(Only ask D3 if the child gives a correct response without explanation/ rationale in D2. If a correct explanation is provided in D2, then give a point in D3 and proceed to D4.)* **Why do you think that the cat is feeling angry/ disappointed/ hurt etc.?** *(use the same IST provided by the child in response to D2)*	Couldn't catch the butterfly/ fell into the bush It hurts to fall into a prickly bush Butterfly escapes/ got away	*Inappropriate/ irrelevant answer*
D4.	**Why does the boy hold the fishing rod in the water?** *(point to picture 5)* (Episode 2: Goal)	Wants to get/ take his ball (back) Wants his ball (back) (In order) to + VERB (get, take) his ball (back/ out)	To play in the water
D5.	**How does the boy feel?** *(point to picture 6)* (IST as reaction)	Good/ fine/ happy/ satisfied/ pleased	Bad/ angry/ mad/ sad
D6.	*(Only ask D6 if the child gives a correct response without explanation/ rationale in D5. If a correct explanation is provided in D5, then give a point in D6 and proceed to D7.)* **Why do you think that the boy is feeling good/ fine/ happy/ satisfied etc.?** *(use the same IST provided by the child in response to D5)*	Has/ got the ball back Could/ was able to + VERB (get, take)	He is smiling/ he looks like that *or other inappropriate answer*
D7.	**Why does the cat grab the fish?** *(point to picture 5)* (Episode 3: Goal)	Decided/ wants to eat/ have/ steal the fish Takes the chance/ opportunity when the boy is not looking Didn't get the butterfly/ Couldn't get/ take the butterfly Cats like fish *(generic meaning)* Fish are tasty/yummy	Wants to play with the fish
D8.	**Imagine that the boy sees the cat. How does the boy feel?** *(point to picture 6)* (IST as reaction)	Bad/ angry/ sad/ mad/ not good	Fine/ good/ happy/ satisfied/ pleased
D9.	*(Only ask D9 if the child gives a correct response without explanation/ rationale in D8. If a correct explanation is provided in D8, then give a point in D9 and proceed to D10.)* **Why do you think that the boy feels bad/ angry/ mad etc.?** *(use the same IST provided by the child in response to D8)*	Cat ate/ is eating/ took/ has taken his fish Boy wanted to eat/ have the fish (himself) It was the boy's fish	Fishing rod is on the ground *or other inappropriate answer*
D10.	**Will the boy be friends with the cat? Why?**	No - *give at least one reason* (cat ate/ stole the fish) *or any other appropriate answer*	Yes/ I don't know/ *other irrelevant answer*

MAIN also contains scripts for all four stories. These are only used when the story is administered as a picture-supported listing comprehension task (Model story or Retelling). The script of the Cat story is given in (1) below. Bolding indicates overt mentions of goals and internal states in the script that are targeted by the comprehension task (D1, D2, D4, D5, D6, D7). This means that a child will hear *some* of the goals and internal states of the story characters being mentioned, if (and only if) MAIN is administered as Model story or Retelling. As the pictures stay in full view, the combined visual and verbal presentation in the Model story and Retelling modes may make it easier for a child to answer the comprehension questions. By contrast, when MAIN is administered as Telling (i.e. story generation followed by comprehension questions), none of the goals or internal states are presented aurally. Rather, the children must infer all goals and internal states by themselves. Since there are four stimuli sets, story comprehension can be compared across different modalities.

(1) The script of the MAIN Cat story in English.
One day there was a playful cat who saw a yellow butterfly sitting on a bush. He leaped forward because he ***wanted to catch it.*** *Meanwhile, a cheerful boy was coming back from fishing with a bucket and a ball in his hands. He looked at the cat chasing the butterfly. The butterfly flew away quickly and the cat fell into the bush. He* ***hurt himself*** *and was very* ***angry****. The boy was so startled that the ball fell out of his hand. When he saw his ball rolling into the water, he cried: "Oh no, there goes my ball!". He was sad and* ***wanted to get his ball back****. Meanwhile, the cat noticed the boy's bucket and* ***thought****: "****I want*** *to grab* ***a fish****." At the same time the boy began pulling his ball out of the water with his fishing rod. He did not notice that the cat had grabbed a fish. In the end, the cat was very pleased to eat such a tasty fish and the boy was* ***happy to have his ball back****.*

4.4 Assessing narrative comprehension with MAIN

The entire procedure of eliciting narrative production and narrative comprehension with MAIN is audio- and/or video-recorded. For comprehension, the child's responses to the questions are then analysed (usually in transcribed format) and scored. This is done with the help of the MAIN scoring sheets that provide examples of correct answers (1 point) and incorrect answers (0 points), as illustrated in Table 1 for the Cat story.

The Revised version of MAIN (2019) contains a relatively comprehensive set of answers to the comprehension questions, based on more than 24,000 responses that were collected after the publication of the original 2012/2013 scoring protocol. It should be acknowledged though that children are so creative that they sometimes will produce new responses not listed in the scoring sheets. Responses are not scored according to whether the child uses the exact wording in the protocol, but according to whether the child's response carries the same meaning and encodes the targeted story component (goal or internal state) or not. To score a point, an answer does not have to be grammatical or lexically correct, but the listener/scorer must be able to understand it. Answers that are incomprehensible, semantically vague, or grammatically so rudimentary that they leave lots of room for interpretation are not awarded points. Purely gestural, nonverbal responses are not awarded points either.

On the one hand, the scorer's task is not difficult, as s/he just has to compare the child's response to the answers in the scoring sheets. But this seemingly easy task can become complicated when children do not answer as anticipated. The scorer has to consider whether the answer captures the targeted goal, internal state and/or rationale. Some children may also give an answer that according to the original MAIN scoring is wrong, though upon further reflection should be considered correct. An example of such an answer was mentioned at the very beginning of this chapter in connection with the Cat story: *How will the boy feel if/when he sees that his fish are being eaten by the cat?* (picture 6). We have come across children who answer this question about an internal state as reaction with a positive emotion (*fine, good, happy, satisfied, pleased*), which at first glance seems out of place, illogical and incorrect. The MAIN protocol also lists such answers as incorrect, see D8 in Table 1. However, when queried why they think that the boy will feel *fine* or *pleased*, some of these children may provide an appropriate rationale for their unexpected answer: One should be kind to others, especially when they are starving; one should not be mad at animals; one should forgive others their transgressions. Far from being wrong, such answers are examples of advanced reasoning and may also reflect a certain ethical viewpoint and a philosophical mind-set. Scoring such creative answers in a way that does the child justice may necessitate multiple rounds of discussion and the drawing-up of more detailed scoring guidelines (e.g. Bohnacker, 2018).

As mentioned earlier, the answers to the ten questions in the MAIN comprehension task are scored one point each, with a maximum of 10 points. Thus, for every child, there is an overall comprehension score of up to 10 points per story. Such scoring is used by all studies in the present volume, so that their raw scores (individual results, group means, ranges, etc.) can straightforwardly compared with each other.[2] Some studies also report overall response accuracies.

2. An exception is the study by Roch and Hržica on Croatian-Italian, who did not ask the tenth comprehension question and therefore only have a maximum of 9 points per story. Gagarina, Topaj, and Sürmeli, in their longitudinal study of Russian-German and Turkish-German children, also report a maximum of only 9 points per story instead of 10, as they did not ask the tenth comprehension question at every data collection point.

Another way of looking at narrative comprehension is to calculate the response accuracies for individual questions, or response accuracies for particular *types* of comprehension questions, i.e. those targeting a particular story component or a particular type of inference, such as a character's goal, a character's internal state, or the rationale/explanation for this internal state. Two thirds of the chapters in the present volume investigate narrative comprehension in this way. Some contributions do not only report response accuracies but study the children's answer patterns in detail and in a more qualitative way, in order to see which types of inferences the children are able to draw.

5. The contributions of the book

This volume contains nine original research chapters. They explore the narrative comprehension of 812 children in 10 different languages (Croatian, Dutch, Finnish, French, German, Greek, Italian, Lebanese Arabic, Swedish, Turkish). Altogether, they cover 2,450 stories and more than 24,000 responses to the MAIN comprehension questions.

All studies except one report results for the Telling mode, i.e. the children answer the MAIN comprehension questions on the basis of the picture series, without having listened to the story first. In addition, five chapters also report results for Model story or Retelling, so that modalities can be compared. All studies except one concentrate on typically developing children. Seven studies are cross-sectional, two longitudinal.

In what follows, we give snapshots of each chapter but do not discuss the findings of the individual contributions. For a synopsis of results, see Section 6.

Fiani, Henry, and Prévost investigate narrative comprehension in 48 Lebanese Arabic-French bilinguals age 4 to 9 in Lebanon, with Baby Birds and Baby Goats stories administered as Telling. Overall narrative comprehension scores in both languages is explored in relation to age, expressive vocabulary, story production, language dominance and a measure of exposure to stories. Performance on different types of comprehension questions (those targeting goals vs those targeting internal states) is also investigated.

Lindgren and Bohnacker report results for 46 German-Swedish (mainly simultaneous) bilingual preschoolers age 4 to 6 in Sweden, with all four MAIN stories, Baby Birds, Baby Goats, Cat and Dog, done as Telling. They explore narrative comprehension in both languages in relation to age and expressive vocabulary and provide a detailed analysis of the types of answers to different types of comprehension questions. Keeping modality constant, task/story effects are investigated as well.

Bohnacker, Öztekin, and Lindgren investigate narrative comprehension in 100 Turkish-Swedish (mainly early sequential) bilinguals age 4 to 7 in Sweden, in both languages, with all four MAIN stories administered as Telling. Keeping modality constant, task/story effects are investigated. Overall narrative comprehension scores are explored in relation to age, receptive and expressive vocabulary in the two languages, home language input and other background factors. Response accuracies for individual questions are compared across tasks and languages, and different types of answers are analysed. Using a qualitative approach, Bohnacker et al. moreover contrastively scrutinise the comprehension questions and the inferences required for the four MAIN stories.

Kunnari and Välimaa compare narrative comprehension in 16 monolingual Finnish and 16 bilingual Swedish-Finnish children age 5 to 6 in Finland and relate the overall comprehension scores to story production. Two different modalities are contrasted: comprehension performance without having heard the story first (Baby Birds/Baby Goats), and comprehension after listening and retelling the story (Cat/Dog). Performance on different types of comprehension questions (goals, internal states, rationale) is investigated as well.

Roch and Hržica study 30 L1 Croatian L2 Italian children age 5 to 6 in a region of Croatia close to the Italian border, with Baby Birds and Baby Goats done as Telling. They explore narrative comprehension in relation to measures of receptive vocabulary, receptive grammar, language input and socio-economic status. They also compare the performance on different types of comprehension questions (goals, internal states).

Blom and Boerma investigate narrative comprehension in 45 monolingual Dutch and 69 L2 Dutch (L1 Tarifit/Berber and L1 Turkish) children age 5 to 7 in the Netherlands, in the majority language Dutch. This is a longitudinal study. Cat and Dog are administered as Model story, Baby Birds and Baby Goats as Telling. Narrative comprehension is investigated in relation to Dutch vocabulary comprehension and language input at home.

Gagarina, Topaj, and Sürmeli is another longitudinal study. They investigate narrative comprehension in 57 early sequential bilinguals (L2 German, L1 Russian and L1 Turkish) from just before age 3 to age 6. Again, Cat and Dog are used for the Model story, Baby Birds and Baby Goats for Telling. Development over time is explored, as are effects of task and length of exposure. Response accuracies for different types of comprehension questions (querying goals vs internal states) are investigated in detail.

Wehmeier reports results for 199 monolingual and 66 simultaneously bilingual German preschoolers (with various home languages) from age 4 to 5 growing up in Germany. Cat is administered as Retelling, Baby Birds as Telling. Narrative comprehension in German is explored in relation to age, gender, language proficiency and nonverbal cognition.

Peristeri, Andreou, Tsimpli, and Durrleman compare 60 monolingual Greek and 60 L2 Greek children (L1 Albanian) age 6 to 8 in Greece. Half the children are typically developing (30 monolingual, 30 bilingual), the other half have a diagnosis of developmental language disorder (30 monolingual, 30 bilingual). Cat and Dog are administered as a listening comprehension task (Retelling). Narrative comprehension in Greek is investigated in relation to age, language exposure, measures of expressive vocabulary, sentence repetition (for the monolinguals), executive functioning, and a theory of mind task (for the DLD groups).

The book concludes with a commentary chapter by *Pearson*.

6. Narrative comprehension outcomes

Here, we do not summarise each of the nine research chapters with their wealth of findings. Rather, we will try to paint a bigger picture, by putting together different pieces of the jigsaw puzzle from the individual chapters, including our own, trying to find common trends. The following is the picture as we see it.

6.1 Steep increase with age initially

The studies in the present volume cover the age range 2;10 to 9;9, and from the cross-sectional results we can see that there is a clear development of narrative comprehension with age, though not necessarily a linear one. There appears to be a stronger, or steeper, increase in overall comprehension scores from around age three to around age 5;6, after which development seems to be less steep. To illustrate what we mean by this, we have compiled the results for typically developing children, both monolinguals and bilinguals, from the individual chapters and plotted the average response accuracies for each age group and language in Figure 3. The diagram only shows results for the 'hardest' condition, i.e. when children have not listened to the story first, for 692 children.[3] The mean scores reported for each age group and language have been transformed into average response accuracies (max 100%). Figure 3 abstracts away from a number of methodological differences between the studies, such as differences in group size, exposure and participant background, so we ask the reader to take this diagram with a pinch of salt.[4] Still, we

3. The remaining 120 children (of 812 in the volume altogether) only did listening comprehension (Retelling).

4. Results for elicitation in other modes (Model story, Retelling) are not plotted in Figure 3, nor are the results for children with developmental language disorder.

believe that by showing the means for story comprehension cross-sectionally by age, an interesting picture emerges: There appears to be a relatively strong increase in comprehension scores from age 3 onwards, an increase that tapers off after age 5;6. Generalising, comprehension goes up from an overall response accuracy of 20% at age 3;6, to 60%–70% around age 5;6. Some increase is in evidence also between 5;6 and 6;6 (70%–90%). After age 6;6 (or 7;0), there does not appear to be much further improvement, at group level. For the participant groups between 5;6 and 9, means are mostly between 70% and 90%. If one were to consider 60%–70% as a measure of being 'acquired', narrative comprehension could be said to be largely acquired by around age five. Of course, the picture is not quite as neat as this generalisation. Readers will notice that in the youngest age groups, the set of bars at 4;6 is unusually high; these are the results for a group of high-SES German-Swedish bilinguals (see Lindgren & Bohnacker, this volume, for discussion). Readers will moreover notice that three sets of bars in Figure 3 are unusually low (at 4;9, 7;1 and 9;0). These bars represent the results for Arabic-French children in Lebanon (see Fiani et al., this volume, for discussion). Whilst the Lebanese children score considerably lower than the other typically developing children of similar age reported in this volume, a clear age development can be seen also for the Lebanese groups (from 4;9 to 7;1 and to 9;0).

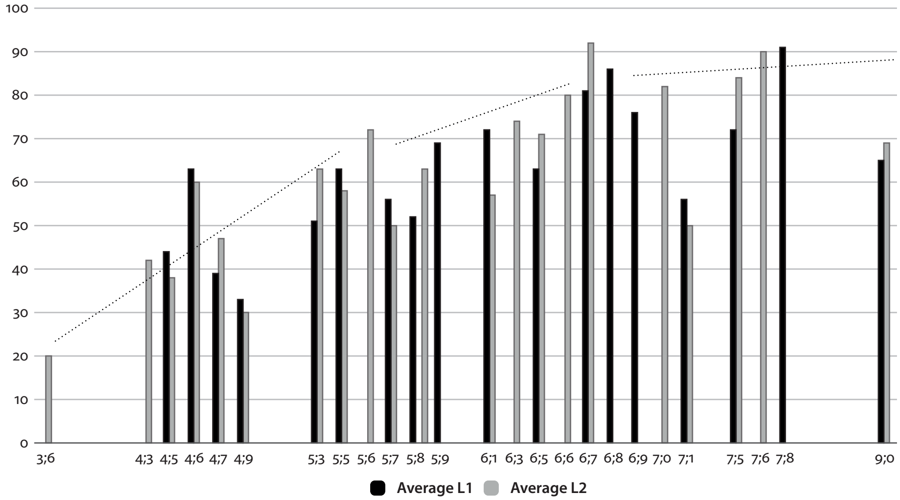

Figure 3. Average overall response accuracies for MAIN comprehension (where max 100 equals 10 questions correct), for each age group and language, compiled from the studies in this book, for the telling modality (i.e. comprehension without having listened to the story first). Black bars show the L1 of bilinguals or monolinguals, grey bars show the L2 of bilinguals. Age in years; months. Dotted lines show the general developmental trajectory.

6.2 Reaching a milestone by age 5

The contributions in this volume show that, for a variety of languages, typically developing children from around age five are generally able to answer the majority (60%–70%) of the comprehension questions correctly. They are able to do so even when they have not listened to the story before. High overall comprehension scores suggest that children are able to infer (and express) most of the goals and internal states of story characters in visually presented stories. Note however that performance is not close to ceiling by age 5 yet (Figure 3).

All contributions that investigate response accuracies for different question types or individual questions find that goals are successfully inferred early on, particularly the more transparent, concrete, goals that concern individual events that occur close together, within an episode (e.g. the boy wanting to get his ball back (out of the water), or the cat wanting to eat the fish). Goals that are more abstract (e.g. the dog wanting to rescue the baby birds) and concern distant events or clusters of events (spanning across episodes) are less successfully inferred. Inferring internal states as a reaction, especially when larger stretches or the entire plotline need to be taken into account (Episode 3), appears to be an even later achievement. This suggests that certain aspects of inferential comprehension are acquired (or mastered) more easily (or earlier) than others (see also Bohnacker, Öztekin, & Lindgren, this volume). Some studies report a task or story effect here, where response accuracies to near-identical comprehension questions are lower (see Section 6.6.).

6.3 Variation at earlier ages

Individual children's narrative comprehension scores vary, but variation is larger for the younger children (age 3 to 5) than for older children.

Irrespective of which language is being investigated, contributions find that substantial numbers of children before age 5 are struggling with the comprehension questions. Even when their language skills are sufficient to engage in meaningful conversation and they are able to answer personal questions (such as during the warm-up for MAIN), many children at this age are unable to answer the comprehension questions correctly. Some young children often just stay within one picture and ignore the plotline when answering the questions, even though all 6 pictures are in full view of the child. These children may have problems recognising that the pictures of the picture series are connected and that content is compositional. They have problems recognising cause-effect links between events and protagonists in the pictures. However, other 3- and 4-year-old children have no trouble answering the majority (or even all) of the questions correctly. Before age 5, the range of comprehension scores is thus large. A low-scoring and a high-scoring four-year-old are contrasted in Table 2 below; both grew up with the same language combination, though the questions and answers are rendered in English here.

Table 2. Cat story: Answers by two four-year-olds

Question	Low-scoring child's answer (age 4;4)	Score	High-scoring child's answer (age 4;11)	Score
D1. *Why does the cat jump forward?*	to catch the fly	1	she wants to get the butterfly	1
D2. *How does the cat feel?*	and then it's here	0	not good	1
D3. *Why do you think that the cat is feeling …?*	–	0	because it really hurts	1
D4. *Why does the boy hold the fishing rod in the water?*	fetch the ball with it	1	cause he wants to get the ball	1
D5. *How does the boy feel?*	not good	0	fine	1
D6. *Why do you think that the boy is feeling …?*	–	0	because he got the ball back	1
D7. *Why does the cat grab the fish?*	yes grab it	0	cause she's hungry	1
D8. *Imagine that the boy sees the cat. How does the boy feel?*	hmm yeah	0	not good	1
D9. *Why do you think that the boy feels …?*	–	0	cause the cat ate the fish	1
D10. *Will the boy be friends with the cat? Why?*	don't know	0	no, because she gobbled up the fish	1
Total (out of 10)		**2**		**10**

At higher ages (age 6, 7 and beyond), children tend to be more homogeneous in their answers, the range of scores shrinks, and overall response accuracies increase. The children often give short, relatively stereotypical, correct answers.

By around age seven, some children begin to give more elaborate and sophisticated answers, especially on the 'what if' scenario discussed earlier (*Imagine that X sees Y. How* …, D8, D9) and the final question, which targets the 'whole plotline' (D10). Such answers reflect higher-order reasoning that is hardly ever evident in the answers of younger children.

6.4 Similarity in narrative comprehension in bilinguals' two languages

Contributions that test bilingual children in both languages find that narrative comprehension scores in the two languages are very similar, at least at group level (see also Figure 3). This appears to be the case even when the children's general language skills are not quite at the same level in their two languages. Interestingly, those contributions that carry out qualitative analyses of the types of answers find very similar types of inferences (both correct and incorrect ones) across languages, and at similar frequencies. All of this suggests that children's inferential narrative comprehension abilities (as measured on MAIN) are not so very dependent on a particular language, and that for bilingual children, narrative comprehension skills transfer between or manifest similarly in their two languages.

6.5 Factors affecting narrative comprehension

All contributions investigate relationships between narrative comprehension and (extra)linguistic factors, in some way or other. Can such factors explain the variation between individual children?

Some contributions explore whether narrative comprehension is affected by independently measured *language skills*, such as receptive and/or expressive vocabulary knowledge, receptive morphosyntactic knowledge, or story production. Here some studies, but not all, find a relationship between story comprehension scores and measures of language proficiency, particularly vocabulary. The mixed results may have to do with the fact that some minimum language proficiency is necessary in order to be able to verbalise the inferred goals and internal states of story characters and communicate them to the experimenter. At very low language levels, the effect on comprehension scores may be strong. However, above a certain language level, there may no longer be such a clear effect on narrative comprehension. As the results do not all go in the same direction, further research is needed here.

Other contributions investigate how story comprehension skills may be influenced by non-linguistic factors, such as chronological *age* and *environmental factors* (such as exposure to narratives, literacy experiences, socio-economic status, language background, exposure, age of onset). All contributions that cover different ages find effects of age, both cross-sectionally and longitudinally (recall also Figure 3), and those that investigate both languages of bilingual children find that story comprehension develops with age in both languages. However, regarding the influence of environmental factors, the picture is much more mixed. Since contributions often explore very different factors or operationalise factors differently, it seems premature to try to make generalisations here.

Finally, two contributions explore child-internal factors, such as nonverbal cognition, executive function or theory of mind, but as they do so for very different participant groups and modalities, it is too early to say how such internal, *cognitive factors* affect story comprehension.

Concerning linguistic and extra-linguistic factors that affect the development of narrative comprehension in children, an emerging field of research opens up before us.

6.6 Putting our tool to the test

The contributions generally report that the MAIN methodology works well: The children seem to like the picture stimuli, they engage in the task, and even at younger ages (age 3 to 5), the comprehension questions elicit a wealth of analysable responses. Apparently, the MAIN method, where the picture series are kept in full view of the child throughout the task, frees up some attentional resources and thus taxes working memory less than other methods, which typically use listening comprehension, with visual support or without, but where the visual support is removed before comprehension is tested.

The contributions do not find floor effects for narrative comprehension in 3- to 4-year-old children, unlike what is often found for narrative production for children this age (e.g. Trabasso et al., 1992; Berman & Slobin, 1994; Hayward et al., 2009). This suggests that MAIN does bring the age down for the analysis of narrative skills in children. At the same time, when combining the results of the few contributions that investigate older children (age 7 to 9), there do not seem to be particularly large gains after age 7, with overall comprehension scores relatively close to ceiling (Figure 3). It could thus be the case that for typically-developing children from around age 7 onwards, MAIN comprehension is 'too easy'. Alternatively, one could focus on how such children perform on the more difficult questions, instead of overall scores (see below). Here, more research is needed.

Contributions that explore how modality affects story comprehension generally find that performance is better when the child first listens to the story (Model story, Retelling) than when answering the comprehension questions without hearing the story first (Telling). Such a task effect is expected. However, the contributions that compare different modalities also use different stories for Model story/Retelling (Cat and/or Dog) than for Telling (Baby Birds and/or Baby Goats). Interestingly, the two contributions that administer all four stories in the *same* modality (Bohnacker et al. and Lindgren & Bohnacker, this volume) find that comprehension scores are higher for Cat/Dog than for Baby Birds/Baby Goats. This suggests that the Model story/Retell advantage is not so much a task effect but likely to be an (unintended) story effect, one that needs to be taken account in future work.

All contributions report overall comprehension scores (of up to 10 points per story), which is very useful for comparisons. Yet overall scores based on answers correct/incorrect (zero vs one point) capture only one aspect of narrative comprehension. Some contributions additionally analyse and classify the different *types of answers* children give, reflecting either a deeper or a more shallow understanding of the story. On a given comprehension question, MAIN awards points for several types of correct answers, so the type of inference is not directly reflected in the response accuracy. A (qualitative) analysis of the types of answers would lead to a more nuanced discrimination of the children's performance. Very few contributions have attempted this, and here more work is needed. Thus another emerging field lies before us, with a more qualitative analysis of picture-based narrative comprehension.

7. Conclusion

If the contributions in this volume are anything to go by, MAIN is indeed working as a tool for narrative comprehension: MAIN offers straightforward comparisons across different ages, languages, children and modalities, it brings the age down for the analysis of children's narrative abilities, and it also offers a variety of analyses (quantitative and qualitative) of narrative comprehension performance in children.

A goal of the volume has been to provide a sense of which narrative comprehension skills children can be expected to master at what age. Here, the combined findings of the contributions point to large gains in comprehension already before the age of 5. Around age 5, a milestone appears to be reached by many children, with good inferential comprehension of the MAIN story character's goals. Bilinguals are reported to have similar narrative comprehension scores in their two languages, a finding that speaks to the question of association of language comprehension in their two languages. The contributions also find that, apart from age, a number of environmental factors affect narrative comprehension, although findings are mixed here, and we are only just embarking on our quest for knowledge. Finding out how such factors may boost or slow down the development of narrative comprehension is important though, as it would allow us to support language development in (bilingual) children in a more informed and better way.

This concludes our introduction to this volume. We now invite the reader to delve into the wealth of detail and results of the individual chapters. We believe that readers will find that the book achieves its goal – to take a large step forward in our understanding of children's narrative comprehension skills. Thus, MAIN as a tool for story comprehension can help us advance our knowledge of language acquisition in general and of language comprehension in particular.

Funding

Funded by: Swedish Research Council.
Award ID: VR 421-2013-1309.
Award recipient: Ute Bohnacker.
Funded by: Open Access Publication Funds of the Leibniz Association.
Award recipient: Natalia Gagarina.
Statement: This work was partly supported by a Swedish Research Council Grant (VR 421-2013-1309) to Ute Bohnacker, by the Open Access Publication Funds of the Leibniz Association to Natalia Gagarina, and by a guest professorship from Uppsala University to Natalia Gagarina.

References

Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137

Astington, J. W., & Pelletier, J. (2005). Theory of Mind, language, and learning in the early years: Developmental origins of school readiness. In B. D. Homer & C. S. Tamis-LeMonda (Eds.), *The development of social cognition and communication* (pp. 312–352). New York, NY: Taylor & Francis (ProQuest Ebook Central).

Berman, R. A., & Slobin, D. I. (1994). Narrative structure. Chapter IIA. In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: A cross-linguistic developmental study* (pp. 39–84). New York, NY: Psychology Press.

Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. Hove: Psychology Press.

Bishop, D. V. M., & Adams, C. (1992). Comprehension problems in children with specific language impairment: literal and inferential meaning. *Journal of Speech and Hearing Research*, 35(1), 119–129. https://doi.org/10.1044/jshr.3501.119

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Bohnacker, U. (2018). *Guidelines for Scoring Macrostructure in MAIN (Unpublished material)*. Uppsala University, Version November 2018.

Bohnacker, U., & Lindgren, J. (in press). MAIN story comprehension: What can we expect of a typically developing child? In S. Armon-Lotem & K. K. Grohmann (Eds.), *LITMUS in action: Cross-comparison studies across Europe*. Amsterdam: John Benjamins. Preprint retrieved from <http://uu.diva-portal.org/smash/get/diva2:1348899/FULLTEXT01.pdf> (18 June, 2019).

Bohnacker, U., Öztekin, B., & Lindgren, J. (this volume). Bilingual Turkish-Swedish children's understanding of MAIN picture sequences: Individual variation, age, language and task effects. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.61.04boh

Boudreau, D. M. (2007). Narrative abilities in children with language impairments. In R. Paul (Ed.), *Language disorders from a developmental perspective: Essays in honour of Robin S. Chapman* (pp. 331–356). Mahwah, NJ: Lawrence Erlbaum Associates.

Burris, S. E., & Brown, D. D. (2014). When all children comprehend: Increasing the external validity of narrative comprehension development research. *Frontiers in Psychology*, 5, Article 168. https://doi.org/10.3389/fpsyg.2014.00168

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: An investigation of their narratives and memory. *International Journal of Language and Communication Disorders*, 43(2), 201–218. https://doi.org/10.1080/13682820701366147

Ellis Weismer, S. (1985). Constructive comprehension abilities exhibited by language-disordered children. *Journal of Speech and Hearing Research*, 28, 175–184. https://doi.org/10.1044/jshr.2802.175

Fiani, R., Henry, G., & Prévost, P. (this volume). Narrative comprehension in Lebanese Arabic-French bilingual children. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins.

Filiatrault-Veilleux, P., Bouchard, C., Trudeau, N., & Desmarais, C. (2015). Inferential comprehension of 3–6 year olds within the context of story grammar: A scoping review. *International Journal of Language & Communication Disorders*, 50(6), 737–749. https://doi.org/10.1111/1460-6984.12175

Florit, E., Roch, M., & Levorato, M. C. (2011). Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. *Discourse Processes*, 48(2), 119–138. https://doi.org/10.1080/0163853X.2010.494244

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). Multilingual Assessment Instrument for Narratives (MAIN). *ZAS Papers in Linguistics*, 56. Berlin: Zentrum für Allgemeine Sprachwissenschaft.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–276). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-011

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics, 63*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, 13(2), 175–197. https://doi.org/10.1016/S0898-5898(01)00061-4

Hayward, D., Schneider, P., & Gillam, R. B. (2009). Age and task-related effects on young children's understanding of a complex picture story. *The Alberta Journal of Educational Research*, 55(1), 54–72.

Johnson, N. S. & Mandler, J. M. (1980). A tale of two structures: Underlying and surface forms in stories. *Poetics*, 9, 51–86. https://doi.org/10.1016/0304-422X(80)90012-1

Kendeou, P., Bohn-Gettler, C., White, M. J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, 31(3), 259–272. https://doi.org/10.1111/j.1467-9817.2008.00370.x

Kintsch, W. (1988). The role of knowledge in discourse comprehension. A constructive-integration model. *Psychological Review*, 95, 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Labov, W. (1972). *Language in the inner city: Studies in the Black English vernacular*. Philadelphia, PA: University of Pennsylvania Press.

Labov, W., & Waletzky, J. (1967). Narrative analysis. In J. Helm (Ed.), *Essays on the verbal and visual arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society* (pp. 12–44). Seattle, WA: University of Washington Press.

Letts, C., & Leinonen, E. (2001). Comprehension of inferential meaning in language-impaired and language normal children. *International Journal of Language and Communication Disorders*, 36(3), 307–328. https://doi.org/10.1080/13682820110045829

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813313

Lindgren, J., & Bohnacker, U. (this volume). Age, language and inferencing: How Swedish-German bilingual preschoolers understand stories. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins.

Lynch, J. S., & van den Broek, P. (2007). Understanding the glue of narrative structure: Children's on- and off-line inferences about characters' goals. *Cognitive Development*, 22, 323–340. https://doi.org/10.1016/j.cogdev.2007.02.002

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology*, 29, 327–365. https://doi.org/10.1080/02702710802165416

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111–151. https://doi.org/10.1016/0010-0285(77)90006-8

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia*, 42, 1414–1434. https://doi.org/10.1016/j.neuropsychologia.2003.12.016

Mayer, M. (1969). *Frog where are you?* New York, NY: Dial Press.

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, 38(1), 36–76. https://doi.org/10.1598/RRQ.38.1.3

Paris, S. G., & Stahl, S. A. (Eds.). (2005). *Children's reading comprehension and assessment*. Mahwah, NJ: Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410612762

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York, NY: Plenum Press. https://doi.org/10.1007/978-1-4757-0608-6

Renfrew, C. E. (1969 [1997]). *Bus Story Test: A test of narrative speech* (4th ed.). Bicester, Oxon: Winslow Press.

Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology*, 27(6), 960–974. https://doi.org/10.1037/0012-1649.27.6.960

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Norwood, NJ: Ablex.

Stein, N. L., & Policastro, M. (1984). The concept of a story: A comparison between children's and teacher's viewpoints. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 113–155). Hillsdale, NJ: Lawrence Erlbaum Associates.

Thompson, J. G., & Myers, N. A. (1985). Inferences and recall at ages four and seven. *Child Development*. 56, 1134–1144. https://doi.org/10.2307/1130228

Tompkins, V., Guo, Y., & Justice, L. M. (2013). Inference generation, story comprehension, and language skills in the preschool years. *Reading & Writing*, 26(3), 403–426. https://doi.org/10.1007/s11145-012-9374-7

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, 15(1), 249–275. https://doi.org/10.1080/01638539209544812

Trabasso, T., Secco, T., & van den Broek, P. W. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Lawrence Erlbaum Associates.

Trabasso, T., Stein, N., Rodkin, P. C., Park Munger, M., & Baughn, C. R. (1992). Knowledge of goals and plans in the on-line narration of events. *Cognitive Development*, 7(2), 133–170. https://doi.org/10.1016/0885-2014(92)90009-G

Van den Broek, P. (1997). Discovering the cement of the universe: The development of event comprehension from childhood to adulthood. In P. W. van den Broek, P. J. Bauer, & T. Bourg (Eds.), *Developmental spans in event comprehension: Bridging fictional and actual events* (pp. 321–342). Mahwah, NJ: Lawrence Erlbaum Associates.

Van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum Associates.

Wenner, J. A. (2004). Preschoolers' comprehension of goal structure in narratives. *Memory*, 12(2), 193–202. https://doi.org/10.1080/09658210244000478

Westby, C. E. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and reading disabilities* (pp. 157–232). Boston, MA: Allyn & Bacon.

Narrative comprehension in Lebanese Arabic-French bilingual children

Rachel Fiani[1], Guillemette Henry[1] and Philippe Prévost[2]
[1]Saint Joseph University of Beirut, Lebanon / [2]University of Tours, Inserm, France

This chapter examines the development of comprehension of macrostructure in narratives by 48 simultaneous bilingual Lebanese Arabic-French children aged 4–9. Fictional storytelling and narrative comprehension tasks were administered in both languages, using the Multilingual Assessment Instrument for Narratives (Gagarina et al., 2012). The comprehension scores were compared across the two languages and analyzed in relation to age, macrostructure production scores, language dominance, expressive vocabulary, and a composite measure of exposure to stories. The results showed significant age effects on comprehension and no differences between languages, irrespectively of language dominance. Significant correlations were found between comprehension and production scores, between comprehension scores in both languages and exposure to stories in French, and between comprehension and expressive vocabulary. The results suggest that story comprehension is invariant across languages, meaning that a bilingual child carries over narrative abilities across languages even when one language is less dominant than the other. Nonetheless, the results also show that exposure to storytelling affects performance on macrostructure comprehension and that assessment of macrostructure comprehension requiring verbal answers necessitates minimal language proficiency.

Keywords: narratives, macrostructure, comprehension, Lebanese Arabic, French

1. Introduction

The number of bilingual children is growing all over the world and particularly in Lebanon, where different types of bilingualism or polyglossia have always existed (Abou, 1962). The linguistic diversity of the country poses a major diagnostic challenge to Lebanese speech and language therapists (SLTs). Most of the tools at their disposal for the diagnosis of developmental language disorder (DLD) are

https://doi.org/10.1075/sibil.61.02fia

in French or English, they do not always correspond to Lebanese linguistic and cultural contexts, and they have been normed on monolingual children. Yet, assessment of linguistic proficiency in both of a child's languages is recommended by professional organizations, such as the American Speech and Hearing Association and the Royal College of Speech and Language Therapists, in order to decrease the risk of misdiagnosis of language impairment (see also Thordardottir, 2015).[1] So far, studies in Lebanon that have aimed at disentangling typically developing bilingual children from children with DLD have targeted the structural aspects of language, such as morphosyntax, and the lexicon (Zebib, Prévost, Tuller, & Henry, to appear). Currently, language professionals, including SLTs, are increasingly recognizing the benefits of storytelling, which is a strong predictor of children's social skills, literacy, and later scholastic achievements (e.g. Dickinson & Tabors, 2001; Gutiérrez-Clellen, 2002; Griffin, Hemphill, Camp, & Wolf, 2004; Oakhill & Cain, 2012; O'Neill, Pearce, & Pick, 2004). To our knowledge no study has aimed to analyze oral narrative production and comprehension in Lebanon. The purpose of this study is to investigate narrative comprehension in relation to production, vocabulary skills, and bilingualism factors in Lebanese Arabic-French bilingual children in order to contribute to a better understanding of children's narrative competence and to facilitate evaluation of oral language in an ecologically valid context.

1.1 Lebanon's multilingual context and the challenges for language assessment

In addition to Lebanese Arabic, which is spoken by over 90% of the population (Leclerc, 2015), and Modern Standard Arabic (a modernized version of classical Arabic mostly used as a written and oral variety in the media, and taught at school), several languages coexist in Lebanon. In particular, French and English are practiced on a large scale by the Lebanese population (Abou, 1962; Haddad, 1997). They have a second language (L2) status, corresponding to the L2 of 45% and 40% of the Lebanese population respectively (Darwiche Jabbour, 2004; Leclerc, 2015). Exposure to the L2 may occur very early (at home with family members and friends), leading to simultaneous bilingualism. At school, exposure to the L2 starts when children begin attending kindergarten, at the age of 3, but children are also exposed to the L2 in daycares and through the media. Furthermore, in Lebanese private schools, 77% of the total teaching time is done in French and English, versus 23% in Modern Standard Arabic (Hoyek, 2004). Finally, many schools introduce French or English as a third language (after Arabic as the first language and either

1. See <https://www.asha.org/ and https://www.rcslt.org/>.

English or French as the second) during the school curriculum, which means that many Lebanese people are trilingual (Shaaban, 1997).

There are multiple challenges arising from multilingualism for clinical practice in Lebanon. In addition to the paucity of appropriate assessment tools, little is known in general about language development in Lebanese Arabic. As a result, clinicians in Lebanon have been relying on qualitative assessment, which increases the risk of misdiagnosis of language impairment. Recently, a standardized tool aiming to screen for language disorders through a general assessment over several language domains was released (ELO-L, Evaluation of Oral Language in Lebanese Children) (Zebib, Henry, Khomsi, Messarra, & Hreich, 2017).[2] In addition, some studies on bilingual language development in Lebanon have been conducted using some of the tools of the LITMUS (Language Impairment Testing in Multilingual Settings) toolbox developed during COST Action IS0804 (*Language impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment*, 2009–2013), including nonword repetition, sentence repetition and cross-linguistic lexical tasks, as well as a questionnaire for parents of bilingual children (PaBiQ) (see Armon-Lotem & de Jong, 2015; Zebib et al., to appear). These studies have focused on the structural aspects of language and on the lexicon, and have not included the production and comprehension of oral narratives. Narrative skills are not targeted by the ELO-L battery either.

1.2 Development of narrative comprehension

Narratives are implicated in social exchanges and provide rich information about the linguistic development of children. They are considered to be an ecologically valid way to assess not only lexical and morphosyntactic capacities (by looking at the linguistic devices required to produce a cohesive story, which is called microstructure), but also children's capacities for constructing narratives (which is called story grammar or macrostructure) (Botting, 2002). Macrostructure, which is the focus of this study, is a central aspect of narrative competence and concerns the narrative's global structure. For Stein & Glenn (1979), a story typically consists of a setting introducing the characters of the story and providing background information, and a collection of episodes. Each episode is considered to contain five elements: an initiating event (which prompts the character to react and take action), an internal response (which includes the protagonist's goal), an attempt (which includes the actions taken by the character to achieve the goal), a consequence (whether or not the goals have been achieved) and a reaction (the character's

2. The ELO-L is the Lebanese adaptation of the French ELO (Khomsi, 2001).

response to the resolution). Understanding the protagonists' goals implies understanding the characters' mindsets, which translates into use of so-called internal state (IS) terms, such as emotional terms (e.g. *happy, feel, angry*), belief terms (e.g. *think, know*), motivational terms (e.g. *want, need*), and experiential terms (e.g. *see, surprised, thirsty*) (see Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2015). In typically developing (TD) monolinguals, development of production of the different macrostructure elements is slow and continues up until age ten (Bishop & Donlan, 2005; van den Broek, 1997). In children with language impairment, macrostructure has been reported to be a vulnerable domain of storytelling (Bishop & Donlan, 2005; Reilly, Losh, Bellugi, & Wulfeck, 2004; but see Norbury & Bishop, 2003). Interestingly, in bilingual contexts, while literature on the development of microstructure in TD children has reported differences between the L2 and the L1, development of macrostructure has been shown to be unrelated to the specifics of a particular language (Akinci, Jisa, & Kern, 2001; Fiestas & Peña, 2004; Pearson, 2002; Uccelli & Páez, 2007 – but see Lindgren, 2018). This supports the idea that macrostructure is universal cross-linguistically, meaning that if narrative abilities, at the level of story structure, have developed during the acquisition process of one language, they should easily transfer to another language. This makes the assessment of macrostructure particularly interesting for the identification of DLD in a bilingual context, since difficulties in that domain may reflect real impairment rather than insufficient exposure to the language (Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016).

So far, studies on the development of macrostructure in bilingual children have mainly focused on production. Little research has investigated comprehension of macrostructure. Yet, story comprehension skills are an important complement to production and should be integrated into the assessment of narratives (Boudreau, 2008). Assessing comprehension provides children with opportunities to demonstrate understanding of macrostructure despite low scores that may occur in production (Gagarina et al., 2015). Importantly, a number of studies show that children's comprehension of narrative structure develops earlier than their ability to produce narratives (see below), which stresses the need for assessing comprehension alongside production (Norbury & Bishop, 2003). Comprehension requires the implication of several skills, such as making inferences, summarizing, and identifying main ideas (Paris & Paris, 2003). These factors, combined with developing cognitive and executive skills, participate in the development of story comprehension. Considering the numerous factors involved in the acquisition of narrative skills, their development is a gradual process that carries on throughout adolescence and into adulthood (Berman & Slobin, 1994).

For bilinguals, it has been shown that narrative comprehension skills increase as a function of age. This has been shown both in studies where only one of the two languages of the child is tested, as in Maviş, Tunçer, & Gagarina's (2016) study of Turkish-German bilinguals who were evaluated in Turkish, and in studies where both languages are examined, as in Kapalkova, Polišenská, Marková, & Fenton (2016) (on Slovak-English bilinguals), Bohnacker (2016) (on Swedish-English bilinguals), and Lindgren (2018) (on Swedish-German and Swedish-Turkish bilinguals). Moreover, performance in comprehension has been reported to be similar in the languages spoken by the child, which suggests that comprehension of macrostructure may not be language specific (Bohnacker, 2016; Kapalkova et al., 2016; Lindgren, 2018; see also Gagarina et al., 2015), which mirrors what has been found for production.

So far, studies that have looked at comprehension of macrostructure in bilingual children, besides the fact that they are not very numerous, have not systematically detailed performance on the different components of story structure. Most of them have reported on global comprehension scores, which may hide different behaviors across those components. The comprehension task of the Multilingual Assessment Instrument for Narratives (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012), which was used in the present study (see section 1.2.1), largely focuses on goals and IS. Bohnacker (2016) did look at comprehension of these two aspects of macrostructure in her study of Swedish-English bilinguals, reporting high scores for goals as of age five. Lindgren's (2018) interpretation of goal understanding is in line with Bohnacker's findings, as the four-year-olds showed good comprehension of goals despite not always including them in production. For IS, scores differed across the different questions of the comprehension task, with lower performance on the question that required taking into account the unfolding of the whole narrative than on those targeting single elements of the story.

Moreover, studies of comprehension so far have been relatively limited in age span, focusing largely on five- to seven-year-old bilinguals (one exception is Maviş et al.'s (2016) study of children aged 2;11 to 7;11, but it focused on one of the children's languages only, i.e. Turkish). Since macrostructure has been shown to be slowly developing, it would be useful to look at comprehension beyond the age of seven, which could also have important clinical implications.

Finally, some studies have looked at the relationship between production and comprehension of macrostructure in bilinguals. Maviş et al. (2016) and Roch, Florit, & Levorato (2016) found significant correlations between the two, although in Roch et al.'s study of five-six year-old English-Italian children, this was observed in all children for English, but only in the six-year-olds for Italian. Kapalkova et al. (2016) found significantly higher scores for production in Slovak compared to English, but no difference between the two languages concerning comprehension, as said above. In Bohnacker (2016) and Lindgren (2018), performance in comprehension was better than in production for all children, in each of their languages.

1.3 Factors involved in the development of narrative comprehension

Although most studies on narrative comprehension by bilinguals found no differences between the child's languages, a language effect was reported in some cases. Roch et al. (2016) found similar performance on narrative comprehension in the two languages of their sequential Italian-English children, but only for six-year-olds. For five-year-olds, comprehension was better in Italian (the children's L1) than in English (the L2) (see also Kapalkova et al., 2016). This differs from the findings reported in studies largely focusing on simultaneous bilinguals (Bohnacker, 2016; Kunnari, Välimaa, & Laukkanen-Nevala, 2016). Moreover, different results have been reported on simultaneous bilinguals depending as to whether the targeted language was the majority or the minority language. Examining Swedish-German and Swedish-Turkish children (age 4–6), Lindgren (2018) found a significant correlation between production and comprehension in Swedish, the majority language, but not in German and Turkish, the minority languages. Related to this issue is the question of whether language dominance has an effect on narrative comprehension in bilingual children. This has not been investigated thoroughly in previous literature, although language dominance would be an important aspect to take into account when comparing results across studies. In her study, Lindgren reported that the Swedish-German children performed better in narrative comprehension (in Swedish) than the Swedish-Turkish children, and that according to parental ratings more children in the Swedish-German group tended to be stronger in Swedish than in their other language compared to the Swedish-Turkish group (Lindgren, 2018, p. 39).

Moreover, although it has been shown that the development of macrostructure in bilinguals is less dependent on the characteristics of the child's languages than for microstructure, the fact remains that being able to answer comprehension questions on the unfolding of a story, its characters and their motivations require minimal language skills (Pearson, 2002). Although, as observed by Bohnacker (2016), what these minimal skills are is not always specified, there is indication that lexical abilities are involved. Westerveld (2014) found significant positive correlations between performance on an expressive vocabulary task and a narrative comprehension task in English-Samoan children aged 4;0 to 4;11, for each of their languages. However, investigation of the role played by vocabulary skills in performance on story comprehension in bilingual children is quite rare.

Another aspect of macrostructure development that has been rarely touched on in previous literature on bilingual children is the impact of exposure to narratives. Shared book reading between caregivers and children expose children to rich narrative experiences and to different types of story structures (e.g. Bus, IJzendoorn, & Pellegrini, 1995; Sparks & Reese, 2013). While a number of studies have shown that the home literacy environment positively affects lexical skills (e.g. Sénéchal & LeFevre, 2002; Sénéchal, 2006), including in contexts of bilingualism (e.g. Gonzalez & Uhing, 2008), few studies have analyzed the effect of exposure to stories on narrative skills, especially on macrostructure comprehension. With respect to production, some studies have reported that the frequency of storytelling has an effect on children's narrative skills in L1 acquisition (Leseman, Scheele, Mayo, & Messer, 2007; van Dongen & Westby, 1986) and in bilingual children (Bitetti & Hammer, 2016). In a study of narratives produced by Lebanese children, which to our knowledge was the first study of its kind in Lebanon, Fiani, Roch, & Henry (2016) looked at production of macrostructure elements in Lebanese Arabic, using Mayer's (1969) *Frog, where are you?* story. They developed a parental questionnaire aimed at measuring children's exposure to stories (see below), which was inspired by the finding that storytelling is not a very common activity in Lebanon, either at home or at school (Lassalle-Gharrios & Marcoin, 2011). Fiani, Roch, and Henry (2016) reported a positive correlation between exposure to narratives and production of macrostructure components in Lebanese Arabic. However, qualitative analyses of the answers to the questionnaire showed that parents preferred telling stories to their children in French rather than in Lebanese Arabic. Among the reasons provided were the facts that books in Arabic are uncommon in libraries in Lebanon compared to the availability of books in French, and that books in Arabic are written in Standard Arabic, which is difficult for children to understand before acquiring written language.

2. Objectives

The present study aims to analyze the developmental characteristics of comprehension of macrostructure, in particular the characters' goals and internal states, by bilingual Lebanese Arabic-French children growing up in Lebanon. The children's narrative skills were examined in both of their languages, with the following objectives:

1. To investigate the development of narrative comprehension in Lebanese Arabic and French.
2. To compare the children's narrative comprehension skills in both languages.
3. To investigate the relationship between comprehension and production of macrostructure elements.
4. To analyze the development of narrative comprehension in relation with language dominance, expressive vocabulary, and exposure to stories.

3. Method

3.1 Participants

Forty-eight Lebanese-French bilingual children, age 4 to 10, participated in the study (19 boys and 29 girls). They were divided into three age groups (4–5, 6–7, and 8–9), henceforth G4–5, G6–7 and G8–9, as shown in Table 1 below. They were recruited from three French medium schools across Beirut. All participants came from a mid socio-economic background based on information on parents' education (collected via the PaBiQ, Tuller, 2015). Most of the mothers in our study had attended primary school and highschool, but not college, and eight of them left school before the age of 15. The children had not been identified as having language difficulties, motor problems, sensory, or neurological disorders. Results from Raven's Progressive Matrices (Raven, Court, & Raven, 1986) confirmed that the children had normal cognitive abilities. The children were also tested in Lebanese Arabic using the ELO-L (Zebib et al., 2017). All children in our population had normal global scores on this test battery. They were therefore considered to be typically developing.

The lexical (expression) subtest of the ELO-L was used as a measure of expressive lexical skills. This probe is a denomination task. Accuracy rates differed significantly across the three age groups, as revealed by Kruskal-Wallis nonparametric tests ($X^2(2, 48) = 36.207$, $p < .0001$, $E^2_R = .770$). Paired comparisons showed that all age groups differed from each other significantly ($p < .0001$).

Data obtained from the PaBiQ on age of onset, language exposure contexts before the age of four, current language skills, languages used at home and languages used outside the home during routine activities were used to calculate a language richness index (LRI) for each language (max. 49 or 50 points).[3] A language dominance index (LDI) was then obtained by subtracting the LRI for French from the LRI for Lebanese Arabic. A positive LDI indicated dominance in Lebanese Arabic, whereas a negative LDI meant that the child was dominant in French. The data obtained from the PaBiQ showed that all children were simultaneous Lebanese Arabic-French bilinguals, except for two children in G6–7 with an age of onset of L2 (French) at 30 and 36 months respectively. Moreover, as shown in Table 1, all age groups displayed a positive LDI. In fact, a large majority of children had a high positive LDI (> 10) in each group, indicating that most children were dominant in Lebanese Arabic: 60% of the children in G4–5 (9/15), 63% in G6–7 (10/16), and 71% in G8–9 (12/17). Only one child, in G4–5, had a high negative LDI (−26.7),

3. The maximum score for the LRI depended on the number of contexts of language exposure before the age of 4 (e.g. presence or absence of siblings in the family).

indicating French dominance. The rest of the children had low LDI, suggesting more balanced bilingualism. The variability in LDI, with children being mostly dominant in Lebanese Arabic, was deemed to be representative of the Lebanese population, since Lebanese Arabic is used every day at home and at school. Kruskal-Wallis nonparametric tests revealed no significant difference across groups for the LDI ($X^2(2, 48) = 3.559$, $p = .169$, $E^2_R = .073$).

The parental questionnaire developed by Fiani et al. (2016) was used to gather information about the children's exposure to narratives. The answers given by the parents allowed us to extract an exposure to narratives index (ENI) for each language (ENI-Lebanese and ENI-French). This index aimed to reflect the wealth of exposure to oral narratives (i.e., stories that are read to the children by their caregivers) before the age of 4 and currently, as well as the wealth of exposure to written narratives (i.e., stories that children read by themselves), in Lebanese Arabic and in French (See Appendix A for the calculation of the ENI).[4] As shown in Table 1, the ENI in French tended to be higher than the ENI in Lebanese Arabic. Wilcoxon nonparametric tests revealed significant differences between the two indexes in G5–6 ($Z(16) = -3.184$, $p = .001$, $r = -.563$) and G8–9 ($Z(17) = -2.300$, $p = .021$, $r = -.394$), but not in G4–5 ($Z(15) = -.932$, $p = .351$, $r = -.170$).

Table 1. Information about the participants: Age, Expressive vocabulary score, Language Dominance Index, and Exposure to Narratives Index for Lebanese Arabic and for French (mean (SD) and range)

	G4–5 (*n* = 15)	**G6–7 (*n* = 16)**	**G8–9 (*n* = 17)**
Age	4;9 (0;3) 4;3–5;3	7;1 (0;6) 6;3–7;8	9;0 (0;7) 8;0–9;9
Expressive Vocabulary	62.6 (4.3) 51.9–68.5	79 (3.4) 71.4–84.3	83.4 (2.6) 80–87.1
Language Dominance Index	13.1 (14.6) −26.7–+27	14.2 (11.4) −8–+29	20.4 (14.2) −4.3–+41
ENI-Lebanese	2.1 (1.3) 0–5	1.8 (1.5) 0–5	2.8 (1.4) 0–5
ENI-French	2.5 (1.4) 0–4	4.3 (1.4) 1–6	3.9 (1.5) 0–6

Note. Performance on expressive vocabulary is reported in terms of accuracy rates; Language Dominance Index ranged from −49/50 (least dominant in Lebanese Arabic) to +49/50 (most dominant in Lebanese Arabic); ENI-Lebanese = Exposure to Narratives Index for Lebanese Arabic (maximum value = 6); ENI-French = Exposure to Narratives Index for French (maximum value = 6)

4. The questionnaire also included information about the parents' reading habits as well as questions related to the types of narratives told to the children. However, these items were not included in the calculation of the ENI.

3.2 Materials

The tool that was used to collect data on production and comprehension of macrostructure was the Multilingual Assessment Instrument for Narratives (MAIN) (Gagarina et al., 2012), which was developed within the framework of COST Action IS0804 as a tool for the assessment of narrative comprehension and production of bilingual children aged 3 to 10 years. MAIN was developed as a cross-cultural assessment tool, which contrasts with the *Frog Story*, which according to Fiani et al. (2016) may be culturally inappropriate for children in Lebanon. Moreover, adopting MAIN allowed for direct comparisons with the growing number of studies on narratives skills in bilingual children based on this tool.

MAIN offers several modes for eliciting narratives from children, including story generation, and retelling and telling after listening to a model story. Story generation was used in the present study, which has been argued to be particularly revealing of the children's spontaneous narrative skills, and more so than other elicitation modes (see Gagarina et al., 2015). MAIN contains four comparable stories (Baby Birds, Baby Goats, Cat, Dog) that are controlled for cognitive and linguistic complexity, as well as for cultural appropriateness (Gagarina et al., 2012). Each story is illustrated by six picture sequences representing three narrative episodes. Each episode contains a goal, an attempt and an outcome. Ten comprehension questions are asked following story production.

The present study is based on the Lebanese Arabic and French versions of MAIN.[5] We used the Baby Birds story for Lebanese Arabic and the Baby Goats story for French. Given that the Arabic language is written from right to left and that books are printed that way too, including children's picture books, the adaptation process included a re-orientation of the pictures of the Baby Birds story. In addition, the instructions and the comprehension questions of the original English version were translated into Lebanese Arabic. Two separate translators performed back-translations to English in order to ensure the accuracy of the translated terms. The Lebanese Arabic and French versions were piloted in 2016–2017 with 18 TD bilingual Lebanese children. A qualitative analysis revealed lexical difficulties in producing the Baby Goats story in Lebanese Arabic, even by children who were dominant in that language. Most of the nouns (especially the ones related to the characters and the setting) and some of the verbs (in particular *drown* and *rescue*) were produced in French, which induced a high level of code-switching and affected the analysis of the results regarding story production in both languages. This

5. The Lebanese Arabic version we used was developed by Rachel Fiani and Guillemette Henry (Higher Institute of Speech and Language Therapy, Saint Joseph University of Beirut, Lebanon) and the French version was developed by Martin Haiden (University of Nantes, Nantes, France) and Alfred Knapp (University of La Rochelle, La Rochelle, France).

led us to use Baby Birds for Lebanese Arabic and Baby Goats for French without counterbalancing the stories across the two languages (however, the order of the languages was balanced – see Section 3.3; see also Section 5 for further discussion). In addition, comprehension questions related to the characters' IS were adjusted in Lebanese Arabic following the results of the pilot study, due to misunderstanding of the word /ʃuˤuːr/ (شعور 'feeling') by the majority of the children. The word was replaced by its synonym /ʔəħseːs/ (إحساس 'feeling') which is considered to be more frequent and to have an earlier age of acquisition.

3.3 Procedures

This research was approved by the Ethical Committee of Saint Joseph University of Beirut, and an informed letter of consent was signed by the parents of the participants. All participants were individually tested in a quiet room at their school, while the parental questionnaires were later administered over the phone. The children completed all the tests mentioned above in two separate sessions each lasting approximately 30 minutes. One of the sessions contained the ELO-L followed by the MAIN story in Lebanese Arabic and the other session contained Raven's Progressive Matrices followed by the MAIN story in French. In order to minimize cross-language influence and carry-over effects, the two languages were assessed with an interval of seven days, and the order of presentation was counterbalanced with regard to language (25 children started with the Lebanese Arabic session and 23 started with the French session).[6] Two different research assistants, native speakers of Lebanese Arabic and French, met with the children, one for each language, in order to establish a monolingual context.

Guidelines for MAIN assessment were strictly respected (Gagarina et al., 2015). Throughout the procedure, the experimenter is not allowed to look at the pictures with the child in order to control the effect of shared knowledge during the presentation of the picture sequences. Three envelopes are placed before the child, who is asked to choose one and to explore the pictures found inside. All envelopes contain the same story, but the child believes that the experimenter is not aware of the content of the story (s)he has 'chosen'. After the child finishes looking at all of the pictures, the experimenter folds them back in pairs without looking at them. When the child finishes relating the events of pictures 1 and 2, the experimenter unfolds pictures 3 and 4, and then pictures 5 and 6 so that all of the pictures are visible to the child at the end. The pictures are presented two at a time in order to facilitate the production of the three episodes in each story. During the storytelling, only minimal prompts such as "tell me more" and "continue" are allowed if the child is silent in

6. No order effect was found in any group for neither language.

the middle of the story. After the child confirms that (s)he has finished telling the story, the experimenter proceeds to ask the comprehension questions, during which the whole story is placed on the table (visible to the child and the experimenter).

All sessions were audio- and video-recorded, including the oral language assessment and the warm-up phase. The stories told by the children and their answers to the comprehension questions were transcribed into CHAT format, following the CHILDES guidelines (MacWhinney, 2000). The transcriptions were then verified by two independent researchers. When opinions regarding transcription and scoring differed, final decisions were reached during research team meetings.

3.4 Narrative measures

3.4.1 *Comprehension*

Out of the 10 comprehension questions asked after the child told a story, three targeted the child's understanding of the main characters' goals for each of the three episodes and six targeted understanding of internal states, and one concerned understanding of the general plotline (see Table 2 for comprehension questions in the Baby Goats story). For the IS questions, three concerned the main characters' internal states (one for each episode) and three questions were follow-ups to these items. They consisted of *why*-questions meant to elicit the reason for the character's IS.

Table 2. Contents of the narrative comprehension task (MAIN)

Story component	Question number	Episode	Examples (from the Baby Birds story)
Goal	D1	1	Why does the mother bird fly away?
	D4	2	Why is the cat climbing the tree?
	D7	3	Why does the dog grab the cat's tail?
IS	D2	1	How do the baby birds feel?
	D3 (follow-up)	1	Why do you think that the baby birds are feeling bad/ hungry etc.?[a]
	D5	2	How does the cat feel?
	D6 (follow-up)	2	Why do you think that the cat is feeling bad/ hungry etc.?[b]
	D8	3	Imagine that the dog sees the birds. How does the dog feel?
	D9 (follow-up)	3	Why do you think that the dog feels good/fine/happy/ satisfied etc.?[c]
Plotline	D10	–	Who does the mother bird like best, the cat or the dog? Why?

Key: IS = Internal State term

a. Use the same IS term provided by the child in response to D2

b. Use the same IS term provided by the child in response to D5.

c. Use the same IS term provided by the child in response to D8.

For each correct answer, the child received 1 point. In the current study, developmental and cross-linguistic analyses were conducted using the general comprehension score (maximum 10 points), the goal-related questions score (maximum 3 points) and the IS-related questions score (maximum 6 points).

3.4.2 *Production*

3.4.2.1 *Story structure*

Story structure refers to the production of macrostructure elements. All studies that were conducted using the MAIN protocol relied on the same story structure score, as explained in the MAIN scoring guidelines. The target macrostructure of the MAIN stories involves a setting and three episodes each containing an initiating event, a goal, an attempt, an outcome and a reaction by one of the main protagonists. The story structure score takes into account the sum of all the macrostructure components of the three episodes of the story told by the child (maximum 17 points).

3.4.2.2 *Structure complexity*

Story complexity refers to the production of different macrostructural sequences (containing attempt-outcome, goal-attempt, goal-outcome, or goal-attempt-outcome sequences). In this study, each narrative generated by a child was assigned a total complexity score based on the production of the different types of sequences in the different episodes. This complexity score was calculated based on the highest level of complexity reached by the child on all three episodes combined (see also Maviş et al., 2016). For each episode, one point was awarded for an attempt-outcome sequence (e.g. for first episode of the Baby Goats story: *The mother goat went into the water and got her baby out)*, two points for goal-attempt (e.g. *the mother goat wanted to rescue her baby so she went into the water)* or goal-outcome sequences (e.g. *the mother goat wanted to rescue her baby and she got it out of the water)*, and three points for full episodes of goal-attempt-outcome sequence (e.g. *the mother goat wanted to rescue her baby so she went into the water and got the baby out)* (maximum 9 points).

4. Results

Since the distribution of the data for the variables of interest was not normal, either within each of the three age groups or in the whole population of our study, as revealed by Shapiro-Wilk tests, nonparametric tests were used for the statistical analysis of the results. Comparisons across the three age groups were performed via Kruskal-Wallis tests (Bonferroni corrections were applied for multiple comparisons) and the relationships between different variables were investigated via Spearman bivariate correlations.[7]

4.1 Development of narrative comprehension in Lebanese Arabic and French

Figure 1 and Figure 2 display accuracy rates on comprehension questions for Lebanese Arabic and French respectively (for raw scores, see Appendix B).

Total comprehension was found to increase with age in both languages. Significant differences were found across the three age groups in each language (Lebanese Arabic: $X^2(2, 48) = 22.391$, $p < .0001$, $E^2_R = .475$; French: $X^2(2, 48) = 20.592$, $p < .0001$, $E^2_R = .420$). For each language, pairwise comparisons revealed significant differences between G4–5 and G6–7, and between G4–5 and G8–9 (with $p < .001$ throughout), but not between G6–7 and G8–9.

Similar developmental patterns were found with respect to comprehension of IS in each language: significant differences across age groups (Lebanese Arabic: $X^2(2, 48) = 21.098$, $p < .0001$, $E^2_R = .431$; French: $X^2(2, 48) = 21.889$, $p < .0001$, $E^2_R = .447$), with pairwise comparisons displaying significant differences between G4–5 and G6–7, and between G4–5 and G8–9 (with $p < .01$ throughout), but not between G6–7 and G8–9. For comprehension of goals, although significant differences were also found across the three age groups in each language (Lebanese Arabic: $X^2(2, 48) = 12.768$, $p = .002$, $E^2_R = .261$; French: $X^2(2, 48) = 9.390$, $p = .009$, $E^2_R = .192$), significant differences only arose between G4–5 and G6–7 in Arabic Lebanese ($p = .003$), while for French significant differences were found between G4–5 and G6–7 ($p = .015$) as well as between G4–5 and G8–9 ($p = .011$), but not between G6–7 and G8–9.

7. We decided to limit statistical analyses to inter-/intra-group comparisons and correlational analyses because of the preliminary nature of our study and our limited sample of 15–17 children per age group. We acknowledge that the line of inquiry adopted here should be pursued both with a larger sample size (given the large number of variables) and the use of more sophisticated statistical tools.

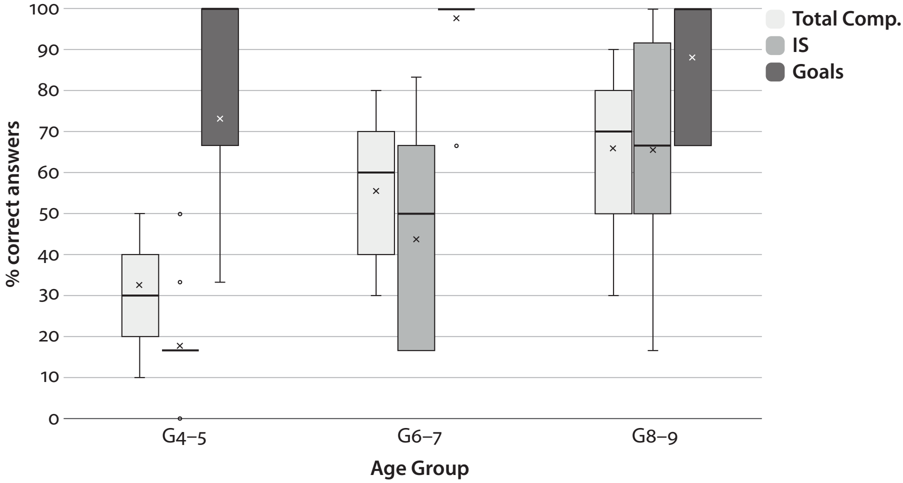

Figure 1. Percentage of correct answers on total comprehension (Total Comp.), comprehension of internal states (IS) and comprehension of goals for Lebanese Arabic in G4–5 (n = 15), G6–7 (n = 16), and G8–9 (n = 17)

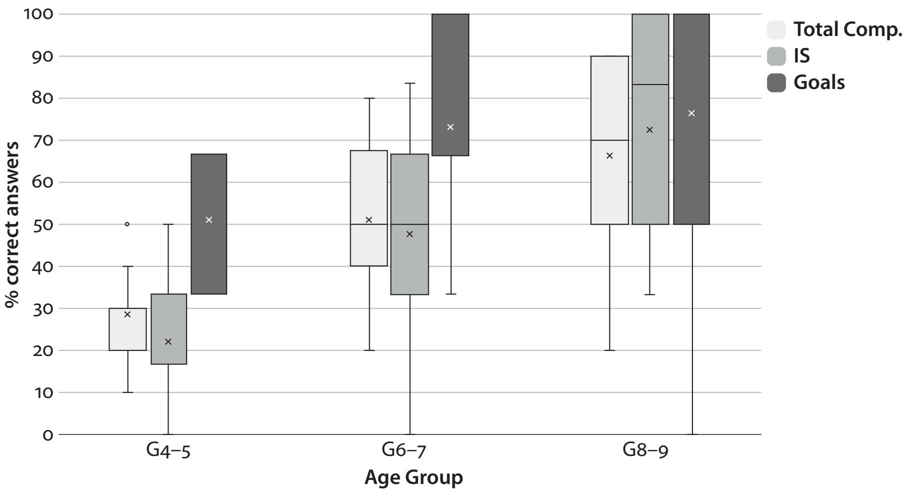

Figure 2. Percentage of correct answers on total comprehension (Total Comp.), comprehension of internal states (IS) and comprehension of goals for French in G4–5 (n = 15), G6–7 (n = 16), and G8–9 (n = 17)

Detailed results for each comprehension question are presented in Table 3.

Table 3. Percentage of correct answers for each comprehension question, in G4–5 (n = 15), G6–7 (n = 16), and G8–9 (n = 17), for each language

Question	Lebanese Arabic			French		
	G4–5	G6–7	G8–9	G4–5	G6–7	G8–9
D1 (Goal)	60	100	100	6.7	56.3	64.7
D4 (Goal)	100	93.8	82.4	93.3	93.8	88.2
D7 (Goal)	53.3	100	82.4	46.7	75	76.5
D2 (IS)	0	31.3	35.3	0	18.8	76.5
D3 (IS; follow-up)	0	0	41.2	0	18.8	52.9
D5 (IS)	20	43.8	76.5	53.3	56.3	88.2
D6 (IS; follow-up)	13.3	43.8	70.6	33.3	37.5	76.5
D8 (IS; ToM)	0	25	41.2	20	43.8	35.3
D9 (IS; follow-up)	0	12.5	47.1	0	37.5	23.5
D10 (plotline)	80	100	100	33.3	81.3	100

As can be seen, comparable patterns of answers in the two languages, apart from the first goal question (D1) and the general plotline question (D10), with much lower accuracy observed for French in comparison to Lebanese Arabic in G4–5. In general, IS-related questions generated lower accuracy than goal-related questions, particularly in G4–5 and in G6–7. The last two questions targeting IS (D8 and D9) were particularly problematic, even for the oldest children (G8–9) for whom accuracy didn't reach 50%.

Finally, significant positive correlations between age and total comprehension, comprehension of IS and comprehension of goals were systematically found for each language, on the whole population (n = 48), as shown in Table 4. As can be seen, age was more strongly correlated with internal states than with goals.

Table 4. Correlations (Spearman's *rho*) between age and total comprehension, comprehension of internal state terms (IS) and comprehension of goals, in each language

	Total	IS	Goals
Lebanese Arabic	.728***	.701***	.363*
French	.670***	.693***	.438**

* $p < .05$
** $p < .01$
*** $p < .001$

4.2 Comparison of narrative comprehension in Lebanese Arabic and French

Comparison of Total comprehension in Lebanese Arabic and French yielded no significant differences in any age group (G4–5: $Z(15) = -1.222$, $p = .222$, $r = -.223$; G6–7: $Z(16) = -1.122$, $p = .262$, $r = -.198$; G8–9: $Z(17) = 0.00$, $p = 1.000$, $r = 0$). No significant intra-group differences were obtained either for comprehension of IS (G4–5: $Z(15) = -1.190$, $p = .234$, $r = -.217$; G6–7: $Z(16) = -.765$, $p = .444$, $r = -.135$; G8–9: $Z(17) = -.672$, $p = .502$, $r = -.115$). For comprehension of goals, significant differences between the two languages arose in G4–5 and G6–7, revealing superior comprehension for Lebanese Arabic over French (G4–5: $Z(15) = -2.500$, $p = .012$, $r = -.456$; G6–7: $Z(16) = -2.972$, $p = .003$, $r = -.525$). There was no significant difference in G8–9 ($Z(17) = -1.513$, $p = .130$, $r = -.259$).

Analyzing the relationship between comprehension in the two languages yielded significant positive correlations for each measure, with weaker coefficient in the case of goals (Total comprehension: $r_s = .699$, $p < .0001$; IS: $r_s = .667$, $p < .0001$; Goals: $r_s = .368$, $p = .01$). Individual analysis revealed that 9/48 children had identical total comprehension scores in both their languages. For 33/48 children, total comprehension scores in the two languages differed by 1 or 2 points; 6/48 children had scores that differed by more than 2 points (at most 5).

4.3 Development of comprehension skills in relation with the development of story structure and story complexity

Figure 3 plots the relationship between story structure and total comprehension scores for each language. Significant positive correlations were found for both Lebanese Arabic ($r_s = .544$, $p < .0001$) and French ($r_s = .658$, $p < .0001$).

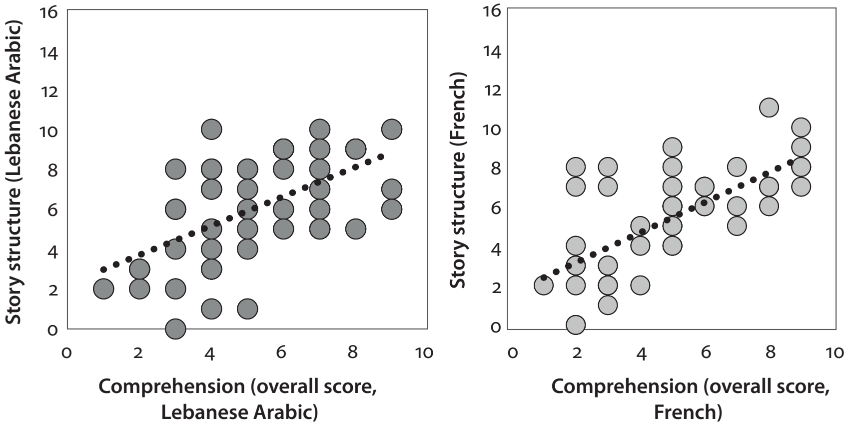

Figure 3. Correlation between production story structure (max 17) and total comprehension scores (max 10) in Lebanese Arabic and in French

Similarly, correlation analyses between structural complexity and total comprehension scores yielded significant results for Lebanese Arabic ($r_s = .370$, $p = .010$) and French ($r_s = .438$, $p = .002$) (see Figure 4).

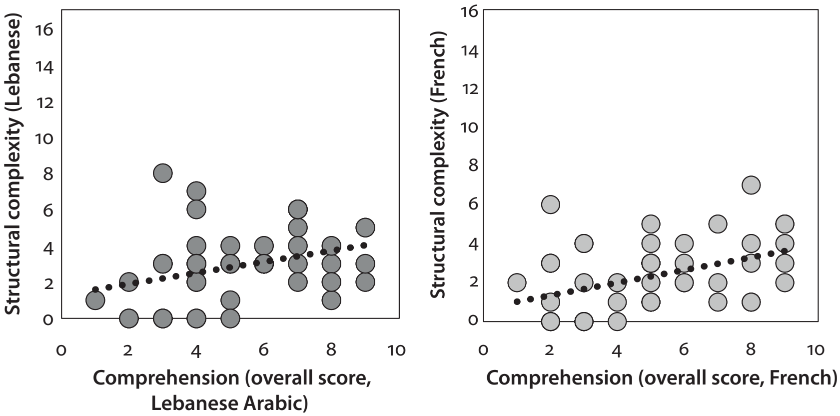

Figure 4. Correlations between structural complexity scores (max 9) and total comprehension scores (max 10) in Lebanese Arabic and in French

4.4 Development of comprehension skills in relation with language dominance and with exposure to stories

As explained in section 3.2.3, an index of language dominance was calculated for every child. Because language dominance varied within each age group, correlations between the LDI and comprehension measures were run for G4–5, G6–7, and G8–9 separately, as well as for the entire group of children. As shown in Table 5, correlation coefficients were generally low across the board, except for G6–7, where the only significant correlation was found (for total comprehension, in French).[8] No significant correlations were found when the whole group of children was taken into account despite the rather high variability in LDI among children. Note that even the child who can be considered an outlier due to his/her strong dominance in French (LDI = −23.75; age = 4;9) scored similarly in both languages on all comprehension measures (5/10 for total comprehension, 2/3 for goal comprehension and 3/6 for IS comprehension).

8. It is not immediately clear why the only significant correlation occurred for total comprehension in French in G6–7. Note that other correlation coefficients larger than −.400 were not associated with low *p*-values (close to .05), which make them difficult to be considered as tendencies. Additional participants would be needed in order to confirm these results.

Table 5. Bivariate correlations (Spearman's *rho*) between the index of language dominance and total comprehension, comprehension of internal states (IS) and comprehension of goals for each language (in each age group)

Age group		**G4–5 (n = 15)**	**G6–7 (n = 16)**	**G8–9 (n = 17)**	**Total**
Lebanese Arabic	Total	−.109	−.434	−.278	*−.018*
	IS	−.105	−.437	−.279	*−.031*
	Goals	.052	−.084	.013	*.043*
French	Total	.158	−.502*	.227	*.206*
	IS	.004	−.410	.210	*.165*
	Goals	.062	−.464	.209	*.146*

* $p < .05$

An exposure to narratives index (ENI) was calculated for each language via information obtained from the parents; see Section 3.2.4. Focusing on the different age groups first, total comprehension, and comprehension of IS and goals were not significantly correlated to the ENI of the corresponding language, in any group (Table 6). However, when the entire group of children was taken together (n = 48), significant positive correlations obtained on the three comprehension measures for French, but not for Lebanese Arabic.

Table 6. Bivariate correlations (Spearman's *rho*) between ENI-Lebanese and total comprehension and comprehension of internal state (IS) and goals in Lebanese Arabic, and between ENI-French and total comprehension and comprehension of IS and goals in French

Age group		**G4–5 (n = 15)**	**G6–7 (n = 16)**	**G8–9 (n = 17)**	**Total**
Lebanese Arabic	Total	.312	.218	.211	*.211*
	IS	.172	.247	.131	*.240*
	Goals	−.100	−.115	.363	*.033*
French	Total	−.016	.419	.010	*.389***
	IS	−.062	.437	−.026	*.380***
	Goals	.159	.380	.121	*.371***

** $p < .01$

When the whole group of children was taken together, we also found that all comprehension measures in Lebanese Arabic were significantly correlated with the ENI in French (total comprehension: r_s = .424, p = .003; comprehension of IS: r_s = .365, p = .011; comprehension of goals: r_s = .346, p = .016). In contrast, comprehension in French was rarely correlated with the ENI in Lebanese Arabic, except for comprehension of IS (total comprehension: $r_s = .261$, $p = .073$; comprehension of IS: $r_s = .352$, $p = .014$; comprehension of goals: $r_s = .006$, $p = .966$). The results on goals were largely due to a significant correlation between the two measures in the youngest group (G4–5, $n = 15$) ($r_s = .575$, $p = .025$). No significant correlations were observed between comprehension of goals in French and the ENI in Lebanese Arabic in the two other groups of children.

Finally, as shown in Table 7, significant positive correlations were found between the raw scores from the expressive vocabulary subtest of the Lebanese Arabic battery ELO-L and the different scores of the comprehension task of MAIN (total comprehension, comprehension of IS and comprehension of goals) in Lebanese Arabic when the whole population was taken into account ($n = 48$). No significant correlations were found within each age group.

Table 7. Bivariate correlations (Spearman's *rho*) between performance (in percentages of correct answers) on expressive vocabulary (ELO-L) and total comprehension, comprehension of internal state (IS) and goals in Lebanese Arabic (in each group)

Age group	G4–5 ($n = 15$)	G6–7 ($n = 16$)	G8–9 ($n = 17$)	Total
Total	.420	.295	.368	*.730*****
IS	.177	.288	.253	*.691*****
Goals	.386	.114	.421	*.428***

** $p < .01$
**** $p < .0001$

5. Discussion

This chapter investigated development of oral narrative comprehension in 48 typically developing Lebanese Arabic-French bilingual children growing up in Lebanon aged 4 to 9. The children, who were all simultaneous bilinguals, were divided into three age groups. Picture-based fictional narratives were elicited in each language, via the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2012). Comprehension questions were then asked of the children, which focused on macrostructure, including the main characters' goals and internal states. Two questionnaires were also administered to the parents (PaBiQ and Exposure to Stories Questionnaire) in order to obtain an index of language dominance and indexes of exposure to narratives in Lebanese Arabic and French.

5.1 Development of narrative comprehension in Lebanese Arabic and French

A significant increase of global comprehension, comprehension of IS, and comprehension of goals was found with age in both Lebanese Arabic and French. These findings are in line with results from other studies using MAIN, which also reported an age effect in comprehension (Maviş et al., 2016). In our study, development of comprehension was found to be particularly strong before the age of 8. Although the comprehension scores of the older group (G8–9) were the highest, no child in that group scored 10/10. This confirms previous findings showing that development of narrative skills is a gradual process that continues throughout adolescence and into adulthood (Berman & Slobin, 1994).

Despite a significant increase with age, the total comprehension scores for our G4–5 and G6–7 groups appear to be lower than the scores of children of the same age reported in other studies using MAIN, such as Bohnacker (2016), Lindgren (2018) and Maviş et al. (2016). One difference, between our study and Bohnacker's (2016) study, concerns the interval between the assessment in the two languages. It was seven days for all of the children in our study, whereas half of the children in Bohnacker (2016) were assessed on the same day, which may have created training effects. Another difference between the two studies has to do with SES, which has been reported to affect children's narrative skills (Bruner, 1990). In Bohnacker's study, the families of the participating children all had high SES, whereas in our study, SES was lower, as measured by parents' education. Similarly, the Swedish-German children in Lindgren's (2018) study all came from high SES backgrounds. Interestingly, the SES profiles of the Swedish-Turkish children in that study were closer to those of our population, and the comprehension scores on Baby Birds/Baby Goats in the Swedish-Turkish group were generally lower than the scores obtained by the Swedish-German children. The scores reported for the Swedish-Turkish group were in fact more comparable to the scores obtained by our Lebanese Arabic-French children than those of the Swedish-German group. Another factor that may explain the differences across studies has to do with exposure to narratives. The indexes that were obtained from our questionnaire on narrative exposure were not very high, especially in Lebanese Arabic. Telling stories to children is in fact not a very common activity in homes and schools in Lebanon (Lassalle-Gharrios & Marcoin, 2011), which might have affected the results in our context. In Bohnacker (2016), regular storytelling and book reading were reported by the parents, and the comprehension skills of their children were rated as being good or very good in both of their languages.

Detailed analysis of the results on narrative comprehension showed that comprehension of IS lagged behind comprehension of goals, especially in G4–5 and

G6–7. These findings are in line with other studies reporting that children before the age of seven still have difficulties fully understanding IS (Bohnacker, 2016). Indication that narratives were generally well comprehended, despite the fact that some questions targeting goals and IS were not answered correctly, comes from the high results on the question related to the plotline (80% in the youngest group).

5.2 Comparison of comprehension skills of children in both languages

No significant differences between Lebanese Arabic and French were found regarding total comprehension and comprehension of IS. Moreover, comprehension scores were close to each other in the two languages. Significant correlations were also found for each measure (total comprehension, comprehension of goals and comprehension of IS). These findings are in line with the few other studies reporting similar comprehension scores in both languages of bilingual children (e.g. Bohnacker, 2016; Kapalkova et al., 2016; Lindgren, 2018). These results support the idea that macrostructure is language independent and may be transferred across languages. Indeed, performance on narrative comprehension involves the person's capacity to draw inferences from a given source, such as a collection of pictures forming a story, as in the case of MAIN. Inferential systems must confront the information provided by the source to the person world's knowledge, and such processing does not narrowly depend on linguistic competence. Moreover, the abstract nature of story structure, which is closely related to discourse representation (see Tsimpli, Peristeri, & Andreou, 2016), would make macrostructure more readily available in the different languages of the child (see also Paris & Paris, 2003; Trabasso & Nickels, 1992). The findings in our study are particularly striking, since this crosslinguistic transfer happens even though the children are little exposed to narratives in Lebanese Arabic.

Significant differences between Lebanese Arabic and French only arose in the two younger groups (G4–5 and G6–7) regarding comprehension of goals. These differences were largely due to the question concerning the goal of the first episode (D1). In G4–5 response accuracy was 60% for Lebanese Arabic and 6.7% for French (see Table 3 above). We contend that this is due to differences in the first episode of the stories. Recall that the Baby Birds story was administered in Lebanese Arabic and that the Baby Goats story was administered in French. The children had little difficulty expressing the first goal in the Baby Birds story (*wants to get food*) presumably because it is based on a more vital and familiar need than the one targeted by the Baby Goats story (*wants to save the baby goat from drowning*). For the latter story, G4–5 children said that they did not know the answer to the D1 question or said *elle veut nager* 'she wants to swim'. In G6–7, response accuracy on this question increased in French (to 56.3%) but it was still lower than the results for Lebanese Arabic (100%). This pattern of results should be compared to the results for the question on the second episodes' goals, which were quite similar in both stories, with a cat and a fox trying to catch a baby bird or a baby goat respectively (question D4). For this question, response accuracy was the highest in both languages in G4–5 (above 93%). Differences across MAIN stories have been reported to affect cross-linguistic comparisons (Lindgren, 2018; see also Gutiérrez-Clellen, 2002, where other narratives were used). In further studies, it would be useful to counterbalance the stories across the children's different languages in order to cancel out potential differences in narratives.

5.3 Development of narrative comprehension in relation with production

Comparisons between story structure (in production) and total comprehension, and between structural complexity (in production) and total comprehension, revealed significant correlations for both languages. Other studies have shown that production and comprehension of narratives are related. Cain (2003) showed that monolingual children who have difficulty understanding narratives also have difficulty producing them. Development of comprehension and production of macrostructure elements are logically linked because they both rely on similar cognitive prerequisites, such as the ability to understand protagonists' intentions and feelings in a story (e.g. Berman & Slobin, 1994; Stein & Glenn, 1979). The development of narrative abilities require advanced theory of mind capacities as well as attention and executive skills (Coelho, Liles, & Duffy, 1995; Oakhill, Hartt, & Samols, 2005).

This notwithstanding, qualitative analysis of our results revealed that some children who answered comprehension questions targeting goals and IS correctly did not include them in their production, especially in G4–5. However, the reverse was rarely observed. This suggests that they had acquired a macrostructure schema that they were able to express in response to prompt questions even though they did not include all components in spontaneous production, which was also shown in other studies (Bohnacker, 2016; Trabasso & Stein, 1994). It could be that producing a whole story, compared to answering specific questions on it, induces more difficulties related presumably to attention, memory and planning skills.

5.4 Development of comprehension skills in relation with language dominance, expressive vocabulary, and exposure to stories

No significant correlations were found across the groups between language dominance and comprehension measures, except for G6–7 (for total comprehension, in French; see footnote 7). No significant correlations were found either when the whole group of children was taken into account. Performance on comprehension thus appears to be unrelated to exposure to one language over another. This result goes hand in hand with the findings in Section 5.2, according to which narrative comprehension is not language dependent, suggesting in turn that underlying processes involved in macrostructure comprehension can be transferred from one language to the other. Individual analysis also revealed no link between comprehension and language dominance.

Other studies reported better narrative production in one language over the other, either the L1 (Kapalkova et al., 2016) or the L2 (Squires, Lugo-Neris, Peña, Bedore, Bohman, & Gilliam, 2014; Uccelli & Páez, 2007). However, these focused on production, and not comprehension. Moreover, the bilingual context of our participants appears to be different from the one of the children in these studies. Our study included simultaneous bilingual children with an early age of onset, in contrast to the studies above, which concerned sequential bilinguals. Moreover, in Lebanon, Lebanese Arabic and French are used daily in society and are granted equal importance in daycares and schools, beginning at age three. In other words, although exposure to L2 French increases upon the child's entry into the school system, this is not detrimental to exposure and use of Lebanese Arabic. This situation differs from the bilingual contexts in Squires et al. (2014) and Uccelli & Páez (2007) where the L2, English, corresponds to the majority language. In these studies, the better performance reported in the L2 over the L1 was attributed to the loss of the L1 during the acquisition of the societal language, which implies that formal schooling in the L2 may influence performance in narrative skills in the L1.

Expressive vocabulary was found to be significantly correlated with performance on comprehension questions, which echoes findings by Westerveld (2014), and is in line with observations by Bohnacker (2016) and Pearson (2002) that although story structure is assumed to be universal, in the sense that it heavily depends on the development of underlying cognitive processes related to the understanding of intentions and motivations of others, its assessment requires minimal language proficiency.

Results from our questionnaire on exposure to narratives revealed a language effect. Narrative comprehension in both French and Lebanese Arabic was significantly correlated with the ENI in French, but not in Lebanese Arabic. Storytelling is known to be related to the development of narrative skills (e.g. Liles, 1993), and some studies also have shown that the number of books at home can affect narrative understanding (Karlsen, Geva, & Lyster, 2016). Our findings for French are in line with these studies, and suggest that exposure to stories plays an important role in narrative comprehension. The reason why the impact of exposure to stories was restricted to French in our study can be explained by the tendency of Lebanese families to tell/read stories to their children in foreign languages, especially French, as showed by Fiani et al. (2016).

Exposure to literacy at school might also have impacted the development of narrative skills. Exposure to written stories has been shown to positively affect comprehension of macrostructure (Dickinson & Smith, 1994). Children in Lebanon start learning written language at approximately five years of age, which can explain the significant increase in both comprehension and production between four and seven years of age, as seen above. Access to literacy at school may in fact act as an accelerating factor in the development of macrostructure skills in Lebanese children given the otherwise low level of storytelling they are exposed to at home, as revealed by Fiani et al. (2016).

6. Conclusion and clinical implications

The findings of this study support the hypothesis of the universality of macrostructure, suggesting that macrostructure is robust to language dominance and paucity of exposure to narratives in both languages. This makes the assessment of narratives particularly appropriate in bilingual contexts, given the difficulties that children with language impairment have with narrative comprehension (Norbury & Bishop, 2003) and production (Bishop & Donlan, 2005; Reilly et al., 2004), including bilinguals (Boerma et al., 2016). Naturally, the first results reported in this study would need to be confirmed in a larger population, with a wider range of language dominance contexts, and it should extend to other language combinations (e.g. Lebanese Arabic-English). Children with language impairment should also be included in these future studies. Finally, the effect of elicitation mode should be investigated. In this study, children were simply asked to tell the story after being shown the pictures, but two other modes could be exploited, such as story retell and story model. In their study of Turkish-German children, Maviş et al. (2016) found that the answers to comprehension questions tended to be more correct when stories were told following a model than when there was no model. The objective of comparing the different eliciting modes would be to come up with the best possible tool for diagnosis.

Acknowledgements

We would like to thank all the children and families who participated in this study. We would also like to thank Rana Nehme, Sabah Bilani, Fatmé Khalife, and Rindala Asmar who participated in the collection of the data. Special thanks are owed to Nour Kassabieh for her significant contribution to our research project on narration, to Lea Eid for her invaluable assistance with data analysis, and to Edith Kouba Hreich, Camille Moitel Messarra, Ute Bohnacker and Natalia Gagarina for their constant support.

References

Abou, S. J. (1962). *Le bilinguisme arabe-français au Liban. Essai d'anthropologie culturelle*. Paris: Presses Universitaires de France.

Akinci, M.-A., Jisa, H., & Kern, S. (2001). Influence of L1 Turkish on L2 French narratives. In L. Verhoeven & S. Strömqvist (Eds.), *Narrative development in a multilingual context* (pp. 189–208). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.23.07aki

Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137

Berman, R. A., & Slobin, D. I. (Eds.). (1994). *Relating events in narrative: a cross-linguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bishop, D. V. M., & Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: Evidence from specific language impairment. *British Journal of Developmental Psychology*, 23(1), 25–46. https://doi.org/10.1348/026151004X20685

Bitetti, D., & Hammer, C. S. (2016). The home literacy environment and the English narrative development of Spanish-English bilingual children. *Journal of Speech, Language, and Hearing Research*, 59(5), 1159–1171. https://doi.org/10.1044/2016_JSLHR-L-15-0064

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

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, 18(1), 1–22. https://doi.org/10.1191/0265659002ct224oa

Boudreau, D. (2008). Narrative abilities. Advances in research and implications for clinical practice. *Topics in Language Disorders*, 28(2), 99–114. https://doi.org/10.1097/01.TLD.0000318932.08807.da

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, 65(1), 1–21. https://doi.org/10.3102/00346543065001001

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology*, 21(3), 335–351. https://doi.org/10.1348/026151003322277739

Coelho, C. A., Liles, B. Z., & Duffy, R. J. (1995). Impairments of discourse abilities and executive functions in traumatically brain-injured adults. *Brain Injury*, 9(5), 471–477. https://doi.org/10.3109/02699059509008206

Darwiche Jabbour, Z. (2004). La francophonie au Liban et les défis de la mondialisation. *Cahiers de l'Association Internationale des Études Françaises*, 56, 17–33. https://doi.org/10.3406/caief.2004.1523

Dickinson, D. K., & Smith, M. W. (1994). Long term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly*, 29(2), 105–120. https://doi.org/10.2307/747807

Dickinson, D., & Tabors, P. (2001). *Beginning literacy with language*. Baltimore, MD: Paul H. Brookes.

Fiani, R., Roch, D., & Henry, G. (2016). Développement de la macrostructure du récit oral: Analyse de la trame narrative chez les enfants libanais âgés entre 5 et 9 ans à partir de l'outil "Frog, where are you" (Unpublished MA thesis). Saint Joseph University of Beirut, Lebanon.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools*, 35(2), 155–168. https://doi.org/10.1044/0161-1461(2004/016)

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Asssessment Instrument for Narratives. *ZAS Papers in Linguistics, 56*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–269). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-011

Gonzalez, J. E., & Uhing, B. M. (2008). Home literacy environments and young Hispanic children's English and Spanish oral language: A communality analysis. *Journal of Early Intervention*, 30(2), 116–139. https://doi.org/10.1177/1053815107313858

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. *First Language*, 24(2), 123–147. https://doi.org/10.1177/0142723704042369

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, 13(2), 175–197. https://doi.org/10.1016/S0898-5898(01)00061-4

Haddad, K. (1997). Anatomie de la francophonie libanaise ou les visages linguistiques du Liban. *Espace, populations, sociétés: Les populations du monde arabe – People of the Arab Middle East*, 1997(1), 65–72. https://doi.org/10.3406/espos.1997.1790

Hoyek, S. (2004). Le français dans l'enseignement scolaire et universitaire au Liban. *Cahiers de l'Association Internationale des Études Françaises*, 56, 49–56. https://doi.org/10.3406/caief.2004.1525

Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak-English children: A cross-language comparison. *Applied Psycholinguistics*, 37(1), 145–164. https://doi.org/10.1017/S0142716415000454

Karlsen, J., Geva, E., & Lyster, S.-A. (2016). Cognitive, linguistic, and contextual factors in Norwegian second language learner's narrative production. *Applied Psycholinguistics*, 37(5), 1117–1145. https://doi.org/10.1017/S014271641500051X

Khomsi, A. (2001). ELO: évaluation du langage oral. Paris, France: ECPA.

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish-Swedish children. *Applied Psycholinguistics*, 37(1), 123–144. https://doi.org/10.1017/S0142716415000442

Lassalle-Gharrios, J. & Marcoin, F. (2011). La rencontre de l'enfant libanais avec le livre. Entre littérature pour la jeunesse française et francophone (Unpublished doctoral dissertation). Université d'Artois, France.

Leclerc, J. (2015). Liban dans l'aménagement linguistique dans le monde. Québec, CEFAN, Université Laval. Retrieved from <http://www.axl.cefan.ulaval.ca/asie/liban.htm> (17 June, 2020).

Leseman, P. P. M., Scheele, A. F., Mayo, A. Y., & Messer, M. H. (2007). Home literacy as a special language environment to prepare children for school. *Zeitschrift Für Erziehungswissenschaft*, 10(3), 334–355. https://doi.org/10.1007/s11618-007-0040-9

Liles, B. Z. (1993). Narrative discourse in children with language disorder and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research*, 36(5), 868–882. https://doi.org/10.1044/jshr.3605.868

Lindgren, J. (2018). *Developing narrative competence. Swedish, Swedish-German and Swedish-Turkish children aged 4–6* (Studia Linguistica Upsaliensia 19). Uppsala: Acta Universitatis Upsaliensis.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish-German bilingual children. *Applied Psycholinguistics*, 37(1), 69–89. https://doi.org/10.1017/S0142716415000429

Mayer, M. (1969). *Frog, where are you?* New York, NY: Dial Press.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, 38(3), 287–313. https://doi.org/10.1080/136820310000108133

Oakhill, J. V., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and writing: An Interdisciplinary Journal*, 18(7–9), 657–686. https://doi.org/10.1007/s11145-005-3355-z

Oakhill, J. V., & Cain, K. (2012). The precursors of reading comprehension and word reading in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 16(2), 91–121. https://doi.org/10.1080/10888438.2010.529219

O'Neill, D. K., Pearce, M. J., & Pick, J. L. (2004). Preschool children's narratives and performance on the Peabody Individualized Achievement Test Revised: Evidence of a relation between early narrative and later mathematical ability. *First Language*, 24(2), 149–183. https://doi.org/10.1177/0142723704043529

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, 38(1), 36–76. https://doi.org/10.1598/RRQ.38.1.3

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853595721-008

Raven, J. C., Court, J. & Raven, J. (1986). *Raven's coloured matrices*. London: H. K. Lewis.

Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language*, 88(2), 229–247. https://doi.org/10.1016/S0093-934X(03)00101-9

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian-English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37(1), 49–67. https://doi.org/10.1017/S0142716415000417

Shaaban, K. A. (1997). Bilingual education in Lebanon. In J. Cummins & D. Corson (Eds.), *Encyclopedia of language and education* (Vol. 5, pp. 251–259). Dordrecht: Kluwer.

Sénéchal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Developmental Psychology*, 73(2), 445–460.

Sénéchal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific Studies of Reading*, 10(1), 59–87. https://doi.org/10.1207/s1532799xssr1001_4

Sparks, A., & Reese, E. (2013). From reminiscing to reading: Home contributions to children's developing language and literacy in low-income families. *First Language*, 33(1), 89–109. https://doi.org/10.1177/0142723711433583

Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gilliam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language and Communication Disorders*, 49(1), 60–74. https://doi.org/10.1111/1460-6984.12044

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives* (pp. 53–120). Norwood, NJ: Ablex.

Thordardottir, E. T. (2015). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 331–358). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-014

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, 15(3), 249–275. https://doi.org/10.1080/01638539209544812

Trabasso, T., & Stein, N. L. (1994). Using goal-plan knowledge to merge the past with the present and the future in narrating events online. In M. M. Haith, J. B. Benson, R. J. Roberts , Jr., & B. F. Pennington (Eds.), *The development of future-oriented processes* (pp. 323–349). Chicago IL: University of Chicago Press.

Tsimpli, I., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment, *Applied Psycholinguistics*, 37(1), 195–216. https://doi.org/10.1017/S0142716415000478

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-013

Uccelli, P., & Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language Speech and Hearing Services in Schools*, 38(3), 225–236. https://doi.org/10.1044/0161-1461(2007/024)

Van den Broek, P. (1997). Discovering the cement of the universe: The development of event comprehension from childhood to adulthood. In P. van den Broek, P. J. Bauer & T. Bourg (Eds.), *Developmental spans in event comprehension and representation* (pp. 321–342). Mahwah, NJ: Lawrence Erlbaum Associates.

Van Dongen, R., & Westby, C. (1986). Building the narrative mode of thought through children's literature. *Topics in Language Disorders*, 7(1), 70–83. https://doi.org/10.1097/00011363-198612000-00009

Westerveld, M. (2014). Emergent literacy performance across two languages: assessing four-year-old bilingual children. *International Journal of Bilingual Education and Bilingualism*, 17(5), 526–543. https://doi.org/10.1080/13670050.2013.835302

Zebib, R., Henry, G., Khomsi, A., Messarra, C., & Hreich, E. (2017). *Batterie d'Evaluation du Langage Oral chez l'enfant libanais (ELO-L)*. Beirut: Liban Tests Editions.

Zebib, R., Prévost, P., Tuller, L. & Henry, G. (to appear). *Plurilinguisme et troubles spécifiques du langage au Liban*. Beyrouth: Presses universitaires de l'Université Saint-Joseph de Beyrouth.

Appendix A. Items included in the Exposure to Narratives Index (ENI)

The Exposure to Narratives Index (ENI) was drawn from parental responses to Fiani et al.'s (2016) questionnaire on children's exposure to narratives. It was calculated by adding the scores obtained on the following items:

- Frequency of exposure to oral stories before the age of 4 (question 5): score of 0, 1 or 2 depending on the frequency of exposure to the story in each language. The score of 0 was awarded if the parents did not tell/read stories to their children before the age of 4. The score of 1 was attributed when the exposure to oral narratives took place at least once a week. The score of 2 was awarded if the parents told stories to their children on a daily basis.
- Frequency of exposure to written stories (question 11): score of 0, 1 or 2 depending on the frequency of exposure to written stories in each language. The score of 0 was awarded if the child did not read stories. The score of 1 was awarded when exposure to written stories happened at least once per week. The score of 2 was awarded if the child read stories on a daily basis.
- Number of books (question 12): score of 0, 1 or 2 depending on the number of books available to the child in each language. The score of 0 was awarded if the child had between 0 and 3 books at home. The score 1 was awarded if the child had between 3 and 10 books at home. The score 2 was awarded if (s)he had more than 10 books at home.

Appendix B.

Mean raw score (*SD*) and range for total comprehension, comprehension of internal states (IS) and comprehension of goals, in Lebanese Arabic and in French

	Lebanese Arabic			French		
	Total (max. 10)	**IS (max. 6)**	**Goals (max. 3)**	**Total (max. 10)**	**IS (max. 6)**	**Goals (max. 3)**
G4–5 (*n* = 15)	3.27 (1.16) (1–5)	1.07 (0.8) (0–3)	2.20 (0.68) (1–3)	2.87 (1.13) (1–5)	1.33 (1.05) (0–3)	1.53 (0.52) (1–2)
G6–7 (*n* = 16)	5.56 (1.46) (3–8)	2.63 (1.36) (1–5)	2.94 (0.25) (2–3)	5.13 (1.86) (2–8)	2.88 (1.54) (0–5)	2.19 (0.66) (1–3)
G8–9 (*n* = 17)	6.59 (1.80) (3–9)	3.94 (1.71) (1–6)	2.65 (0.49) (2–3)	6.65 (2.23) (2–9)	4.35 (1.46) (2–6)	2.29 (0.99) (0–3)

Inferential comprehension, age and language

How German-Swedish bilingual preschoolers understand picture-based stories

Josefin Lindgren[1,2] and Ute Bohnacker[1]
[1]Uppsala University / [2]Leibniz-Centre General Linguistics (ZAS)

This study investigates story comprehension in 46 German-Swedish 4- to 6-year-old bilinguals growing up in Sweden. The children's inferential understanding of goals and emotions of story characters in visually presented stories was assessed in both German and Swedish, using the comprehension questions from the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015) for the narrative tasks Cat/Dog and Baby Birds/Baby Goats. We analysed effects of age, language, and narrative task on overall comprehension scores and investigated whether comprehension scores were influenced by expressive vocabulary knowledge, operationalized as scores on a vocabulary task (Cross-Linguistic Lexical Task, CLT; Haman et al., 2015). Additionally, response patterns for the different comprehension questions were analysed.

We found effects of age, with 6-year-olds outperforming 4- and 5-year-olds, but no significant difference between the two younger groups. The development with age was similar in both languages and was consistent across tasks. The main effect of language was not significant, but when German was tested first, the children performed lower in German than in Swedish. When Swedish was tested first, no difference was found between the languages. The effect of expressive vocabulary was not the same in the two languages. In German, but not in Swedish, CLT expressive vocabulary scores significantly predicted narrative comprehension scores. The children's inferential comprehension performance depended on the narrative task used, with higher scores for MAIN Cat/Dog than Baby Birds/Baby Goats, and response accuracy was also found to vary substantially between different comprehension questions. Response patterns to individual questions were strikingly similar in Swedish and German, suggesting that they may generalize across languages. The results indicate that an analysis of individual comprehension questions allows us to explore and detect patterns not visible in overall scores.

Keywords: German, inferencing, narrative comprehension, Swedish, vocabulary

https://doi.org/10.1075/sibil.61.03lin

1. Introduction

When we comprehend stories, we seek to find relationships between events and create mental representations of the overarching narrative structure, the narrative macrostructure (e.g. Graesser, Singer, & Trabasso, 1994; Kintsch, 1988; Stein & Glenn, 1979; Trabasso, Secco, & van den Broek, 1984). These mental representations are based both on information provided in the stories and on our background knowledge. Story content can be presented in various ways, verbally (a story told), visually (shown in pictures or in silent video-clips) or both verbally and visually (films or stories told accompanied by picture sequences) (e.g. van den Broek, Kendeou, Kremer, Lynch, White, & Lorch, 2005; Paris & Paris, 2003). Neither visually presented nor verbally presented stories make all aspects of a story explicit. Story comprehension thus always necessitates *inferencing*, for example about the goals and emotions of story characters (Bishop, 1997; Letts & Leinonen, 2001; Hayward, Schneider, & Gillam, 2009; Stein & Glenn, 1979). To draw inferences means to integrate various types of information that may be taken from different sources into a conclusion (Graesser et al., 1994, pp. 373–376; Levinson, 1983, pp. 21–22). Our background knowledge includes knowledge of facts of the world, how social interaction works, cause-effect relationships, and knowledge of narrative schemata.

Comprehending a story means understanding the relations between events, including the reasons for a character's actions, i.e. understanding the plot of a story (van den Broek et al., 2005, p. 118, 126). Picture-based narrative comprehension tasks can thus tell us about the child's ability to draw inferences from the pictures and verbalize them in a comprehensible manner when probed, for example concerning story characters' goals and emotions (Burris & Brown, 2014), which in turn requires Theory of Mind, i.e. the ability to understand what the story characters think and feel (e.g. Astington & Pelletier, 2005; Pelletier & Astington, 2004). Comprehension of an underlying story schema may be a prerequisite for being able to tell a well-formed story (Burris & Brown, 2014; Stein & Glenn, 1979; Trabasso & Rodkin, 1994). Results from previous studies of narrative comprehension indicate that "analyzing only what the speaker says may underestimate what the speaker knows" (Trabasso & Rodkin, 1994, p. 103), i.e. that speakers, and children in particular, are able to understand more of the narrative structure than they are able to spontaneously include in their narratives (e.g. Bohnacker, 2016; Lindgren, 2019; Trabasso & Rodkin, 1994). To fully assess children's narrative abilities, it is thus necessary to assess their story comprehension. However, the majority of existing studies have focused on children's narrative production; this is especially the case for studies of bilingual children. This chapter reports a study of 4–6-year-old German-Swedish bilinguals, an understudied group acquiring two closely-related languages, and their narrative comprehension in both languages.

In the present study, comprehension questions (see Section 4.2 for details) from the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker & Walters, 2012, 2015) were used to probe the children's inferential comprehension. MAIN is a battery of picture-based narrative tasks that are suitable for children aged 4–10 speaking a variety of languages. Several studies that have used MAIN to investigate bilingual children's narrative abilities have been published in recent years, including a special issue of *Applied Psycholinguistics* in 2016 (e.g. Bohnacker, 2016; Gagarina, 2016; Kunnari, Välimaa, & Laukkanen-Nevala, 2016; Roch, Florit, & Levorato, 2016). With some exceptions these studies primarily explore children's narrative production, not narrative comprehension. Much is still unknown about bilingual children's narrative skills, especially regarding (inferential) story comprehension, both in the context of MAIN and more generally.

The paper proceeds as follows. Section 2 summarizes findings from earlier studies of children's narrative comprehension. Section 3 states our aim and research questions, and Section 4 describes the methodology. Results are presented in Section 5. Finally, Section 6 contains a discussion and conclusion.

2. Background

In this section, we first summarize general findings of children's comprehension of narratives (Section 2.1), before zooming in on results from studies using MAIN, the same instrument as in the present study (Section 2.2).

2.1 General findings

The development of children's narrative comprehension is less well studied than their narrative production skills (storytelling and retelling). Narrative comprehension has also not always been clearly separated from narrative production; sometimes narrative production tasks such as narrative recall or retell are listed as (indirect) methods for investigating narrative comprehension (e.g. Burris & Brown, 2014). However, when such methods are used, the 'comprehension' measured is in fact the child's ability to remember as many aspects of the story as s/he has been told (Boudreau, 2007; Dodwell & Bavin, 2008), and not necessarily his/her ability to understand an underlying story schema. Such methods thus tax the child's (verbal) short term memory to a large extent, and make it difficult to separate effects of memory from the child's actual comprehension.

When narrative comprehension has been studied directly, it has mostly been in the form of probe questions (e.g. Bishop & Adams, 1992; Bohnacker, 2016; Gibson, Peña, & Bedore, 2018; Letts & Leinonen, 2001; Lindgren, 2018; Stein & Glenn, 1979; Trabasso, Stein, Rodkin, Munger, & Baughn, 1992). Depending on the format in which the story content is presented, the child is asked the probe questions either after s/he has listened to (and in some cases also has retold) a story, or after s/he has looked at pictures and then told a story (as in the method employed in the present study). Compared to retelling a story, answering comprehension questions is less taxing for the child's memory, especially if the picture stimuli are visible throughout, as in the current study, and are not removed, as is often done elsewhere.[1] Answering such questions usually requires inferencing on the part of the child, irrespective of how the story content is presented (orally, visually, or both), and also requires that children formulate their answers verbally.[2] Results indicate that children's comprehension of narrative structure is relatively well-developed at a time when their narrative production is still rudimentary (e.g. Stein & Glenn, 1979, Trabasso et al., 1992). In a well-known study, Stein & Glenn (1979) analyzed answers to (mainly inference-requiring) comprehension questions by monolingual English 6- and 10-year-olds ($N = 24$). Their results showed that the 6-year-olds performed well on inferential comprehension of narratives, with good comprehension of goals and internal states, components which were only rarely expressed when the same children retold stories. Similarly, Trabasso et al. (1992) found that English monolingual 4-year-olds who did not produce story characters' goals in their narratives were able to explain why characters performed certain actions when explicitly probed.

Some previous studies have investigated children's inferential comprehension of narratives with different *types* of comprehension questions (e.g. Bishop & Adams, 1992; Gutiérrez-Clellen, 2002; Letts & Leinonen, 2001; Paris & Paris, 2003), for example on 'factual' vs 'inferential' questions. Here, factual narrative comprehension questions target recall of factual details that were either presented verbally (listening comprehension) or visually (stimulus pictures/film). Inferential questions target

1. When the child is asked to answer comprehension questions *without the pictures present* (e.g. Bishop & Adams, 1992; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012; Letts & Leinonen, 2001; Lynch, van den Broek, Kremer, Kendeou, White, & Lorch, 2008; Omanson, Warren, & Trabasso, 1978; Paris & Paris, 2003; Stein & Glenn, 1979; Trabasso et al., 1984; van den Broek et al., 2005; and the studies in a meta-analysis by Filiatrault-Veilleux, Bouchardt, Trudeau, & Desmarais, 2015), demands on the child's memory are also high, just as is the case in retelling tasks.

2. Such an assessment of narrative comprehension is therefore not independent of language skill. However, responding to comprehension questions does not require as complex expressive language as actual story production does. When story telling or retelling is employed to (more indirectly) investigate narrative comprehension, linguistic demands on the child are higher.

comprehension of story aspects that were not explicitly mentioned or depicted and must be induced/inferred. However, few studies of visually presented stories have explored answers to specific inferential questions of different types in detail, and even fewer studies analyse children's incorrect answers to such questions (pace Bishop & Adams, 1992). The current study investigates different types of answers given by children to specific comprehension questions, with a view to exploring general patterns and potential developments that may otherwise stay 'hidden' in the overall comprehension scores.

As mentioned above, comprehension questions constitute a relatively direct way of measuring narrative comprehension (cf. Liles, 1993), even though answers still need to be expressed verbally and are thus not independent of language skills. It may well be the case that a child fully understands a story but cannot fully convey this understanding due to limited language proficiency. To explore this possibility further, the present study includes an analysis of how expressive vocabulary knowledge relates to narrative comprehension (scores). We already have clear indications from narrative *production* that narrative ability is not independent of general language proficiency, for example vocabulary knowledge (e.g. Iluz-Cohen & Walters, 2012; Lindgren, 2018; Mäkinen, 2014; Tompkins, Guo, & Justice, 2013).[3] For example, in a study of monolingual Swedish and bilingual German-Swedish and Turkish-Swedish children aged 4–6 ($N = 166$) telling MAIN, Lindgren (2018) found a clear effect of expressive vocabulary on children's production of narrative macrostructure, both in the majority language Swedish and in the minority languages German and Turkish.[4] However, the potential link between expressive vocabulary and narrative *comprehension* was not explored in Lindgren (2018). A number of studies of monolingual children have found correlations between receptive or expressive vocabulary knowledge and narrative listening comprehension (e.g. Lepola et al., 2012; Lynch et al., 2008; Westerveld, Gillon & Boyd, 2012). However, to our knowledge, the role of general language proficiency, e.g. vocabulary knowledge, has not been systematically investigated for bilinguals' narrative comprehension.

In sum, previous studies have found children's narrative comprehension to develop earlier than their production. The effects of general language proficiency on bilingual children's narrative comprehension as well as children's performance on different types of inferential questions are as yet unknown. Next, we take a closer look at results from studies using MAIN to assess narrative comprehension.

---

3. Note that vocabulary knowledge is only one aspect of general language proficiency: in order to fully assess children's language, other aspects, including for example syntactic complexity and morphological accuracy, would need to be taken into account as well.

4. Note that the German-Swedish bilinguals of Lindgren (2018) were the same as the children in the present study.

2.2 Comprehension in MAIN

A number of recent studies report overall scores on the comprehension questions from MAIN (see Section 4.2) (Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Bohnacker, 2016; Bohnacker & Lindgren, in press; Kapalková, Polišenská, Marková, & Fenton, 2016; Lindgren, 2018, 2019; Maviş, Tunçer, & Gagarina, 2016; Otwinowska, Mieszkowska, Białecka-Pikul, Opacki, & Haman, 2018; Roch et al., 2016). None of these studies investigate the effects of general vocabulary knowledge on narrative comprehension, but primarily focus on age effects, differences between mono- and bilinguals and/or differences between bilinguals' languages. Performance on specific (types of) comprehension questions is only rarely explored, and only one study (Bohnacker & Lindgren, in press) analyzes children's incorrect answers qualitatively.

In a study of 52 English-Swedish bilinguals aged 5–7 growing up in Sweden who told two parallel stories from MAIN, Baby Birds/Baby Goats (for details, see Section 4.2), Bohnacker (2016) found no significant differences between the bilinguals' two languages in narrative comprehension. In both Swedish and English, the 6–7-year-olds performed better than the 5-year-olds, although comprehension was at a high level already at age 5.

In a large-scale study, Boerma et al. (2016) investigated Dutch narrative production and comprehension in mono- and bilingual Dutch-speaking children with typical development and with language impairment aged 4;10–7;0 (*N* = 132). They found a clear effect of language impairment, but no difference between mono- and bilinguals on Baby Birds/Baby Goats narrative comprehension in Dutch. Age effects were not investigated.

Roch et al. (2016) analyzed both languages of Italian-English bilinguals aged 5–7 (*N* = 62). The bilinguals were growing up in Italy and had attended an English-medium school from around age 3. Children aged 5–6 (who had not yet had that much exposure to English) scored higher in L1 Italian than in L2 English, whereas children aged 6–7 had similar comprehension scores for Baby Birds/Baby Goats in the two languages. For L2 English, there was a large difference between the younger and the older group, but in L1 Italian, there was no significant difference between the age groups. These results indicate that not only age, but also length of exposure and most likely also language proficiency (the younger children in the study had significantly lower scores on the PPVT in English than in Italian) play a role for the children's narrative comprehension.

Bohnacker and Lindgren (in press) investigated story comprehension in 72 Swedish monolinguals aged 4–6 and the 52 English-Swedish bilinguals aged 5–7 from Bohnacker (2016). The monolinguals were tested with MAIN Baby Birds/Baby Goats and Cat/Dog, another set of two parallel stories from MAIN (see

Section 4.2), and were found to perform significantly better on Cat/Dog, irrespective of age group. Bohnacker and Lindgren (in press) found no significant differences between the mono- and the bilinguals, neither for the 5-year-olds, nor for the 6–7-year-olds. In addition to the quantitative analysis of overall comprehension scores, they carried out a detailed investigation of response accuracy and types of answers to specific comprehension questions. To our knowledge, their study is the only existing study that has carried out such an analysis for MAIN comprehension. On the basis of their results, Bohnacker and Lindgren made explicit suggestions for what we can expect of a typically-developing child aged 4–7 in (MAIN) narrative comprehension.[5] First, overall inferential comprehension on Cat/Dog can be expected to be high already at age 4, and by age 6, it should be close to ceiling, whereas accuracy on Baby Birds/Baby Goats cannot be expected to be equally high. Second, for both tasks, by age 6, comprehension of goals should have reached a very high level, close to ceiling. Finally, Bohnacker and Lindgren (in press) found that response accuracy on questions probing the child's inferential comprehension of a hypothetical scenario was especially low for Baby Birds/Baby Goats, but substantially higher on Cat/Dog, and predicted that such differential performance (i.e. a task effect) should be found more generally, and also for children of other language combinations. The present study investigates whether these expectations are also borne out for German-Swedish 4- to 6-year-olds.

To summarize, the results of previous studies using MAIN point to an age development in narrative comprehension (Bohnacker, 2016; Bohnacker & Lindgren, in press; Lindgren, 2018, 2019; Maviş et al., 2016; Roch et al., 2016) and indicate that bilinguals tend to score similarly in their two languages in comprehension, at least at age 6–7 (Bohnacker, 2016; Kapalková et al., 2016; Lindgren, 2018; Roch et al., 2016). There are also indications that (some) bilinguals may score similarly to monolinguals on narrative comprehension (Boerma et al., 2016; Lindgren, 2018). Additionally, children's performance on narrative comprehension may vary between tasks and/or individual comprehension questions (Bohnacker & Lindgren, in press; Bohnacker, Öztekin, & Lindgren, this volume). The relationship between comprehension in MAIN and aspects of general language proficiency such as vocabulary knowledge remains unknown.

5. It should be said that Bohnacker & Lindgren's (in press) suggestions for benchmarks for MAIN comprehension were made on the basis of monolingual and bilingual children from mid- to high-SES backgrounds in Sweden, who were regularly exposed to storytelling and book-reading activities.

3. Aim and research questions

The aim of the present study is to investigate German-Swedish bilingual 4- to 6-year-olds' narrative comprehension in both languages using two different narrative tasks (MAIN Cat/Dog and Baby Birds/Baby Goats), analyzing both overall comprehension scores and response patterns to specific questions. The following research questions are asked (the first four questions concern the children's overall comprehension scores, and the final question is addressed in our analysis of answers to specific comprehension questions):

1. Does narrative comprehension develop with age from 4 to 6 years? Is the effect of age the same in both languages?
2. Is there a difference between the languages (German, Swedish) in narrative comprehension?
3. Do the children perform differently on the narrative tasks of MAIN (Cat/Dog vs Baby Birds/Baby Goats)?
4. Is performance on narrative comprehension in the two languages influenced by general expressive vocabulary knowledge in each language (when age is controlled for)?
5. What are the children's response patterns for specific comprehension questions?

4. Methods

4.1 Participants

The participants were 46 German-Swedish bilinguals aged 4;0–6,11 growing up in Sweden. Informed parental consent was obtained in writing, and the children could terminate their participation in the study at any time. The children were divided into three age groups: 4-year-olds, 5-year-olds, and 6-year-olds. Table 1 gives an overview of the children in the different groups.

Table 1. Overview of the participants

	4-year-olds	5-year-olds	6-year-olds	Total
N (girls/boys)	14 (9/5)	16 (12/4)	16 (10/6)	46 (31/15)
Mean age	4;6	5;5	6;7	5;7
Age range	4;0–4;11	5;0–5;11	6;1–6;11	4;0–6;11

The parents filled in a detailed questionnaire about the child's language development, patterns of language use in the family, and the parents' language and educational backgrounds. All children were living in Sweden at the time of data collection and most (74%, 34 children) had done so from birth. They came from mid- to high-SES backgrounds; all parents had at least some tertiary education, which can be considered typical of German-speaking families in Sweden. The children did not have any diagnosed language impairment according to parental report. They were able to speak both Swedish and German well enough to complete the narrative tasks in each language. All children were continuously exposed to the minority language German from birth in the home. Most children (69.6%, 32 children) were reported to have received regular Swedish input from birth or before age 1;0. For 10 children (21.7%) regular exposure to Swedish started between age 1;0 and 2;0. Only for four children (8.7%) did regular exposure to the majority language Swedish start after age 2.

The children attended institutional childcare for a major part of the day, as is typical in the Swedish context: 30 children (65.2%) attended Swedish-medium (pre)schools, and 14 children (30.4%) attended a Swedish-German bilingual (pre) school. The children were mainly preliterate. Most 4- and 5-year-olds attended preschools, and a few 5-year-olds and most 6-year-olds attended *förskoleklass* 'preschool class', a non-mandatory preparatory year before Grade 1. A few of the older 6-year-olds had started Grade 1 at the time of testing: these were the only children who had any literacy (beyond being able to write their own names and a few words), as reading and writing is not taught in Sweden before age 7, in Grade 1 of primary school.

According to parental report, no child received more than 60% daily input in German. Since the children grow up with Swedish as majority language and attend (pre)schools exclusively or partly in Swedish for a major part of the day, higher levels of exposure to Swedish are expected. The parents of 26 children (56.5%) reported that their child received at least 40% daily input in both languages, whereas 17 children (37%) were reported to receive 80% or more daily input in Swedish. The remaining three children, and one child with 80% Swedish input, were also exposed to a third language.

Slightly over half the children (54.3%, 25 children) had one parent who was a Swedish native speaker and one parent who was a German native speaker. Eighteen children (39.1%) had two parents who were native speakers of German, and two children had one native German parent and one parent with another native language. One child had parents who were both German-Swedish bilinguals themselves.

The parents of almost half of the children (48%) rated their children's expressive language skills to be somewhat higher in Swedish. Only a smaller group (17%) rated their children to be more proficient in German. The remaining children (35%) were rated to have equal expressive skills in both languages.[6]

6. For more details about the participants, see Lindgren (2018: Chapter 3).

4.2 Materials

The present study focuses on the children's comprehension results on MAIN (Gagarina et al., 2012, 2015). MAIN contains four different stories (Cat, Dog, Baby Birds, Baby Goats), of which two each are strictly parallel in plotline (Cat/Dog; Baby Birds/Baby Goats). Each story is a picture sequence with six pictures showing a story consisting of three episodes. Each story episode contains a goal-attempt-outcome sequence (cf. Gagarina et al., 2012). MAIN has a specific procedure for administering the task in the 'telling mode', where the child first looks at all six pictures at his/her own pace before telling the story and is then asked ten standardized comprehension questions (see below). It is this procedure that was used in the present study for every story. The four stories are parallel in terms of length and story grammar components. They differ in the number of story characters: Cat/Dog has three characters and Baby Birds/Baby Goats five characters. In what follows, Cat/Dog will be referred to as one narrative task, and Baby Birds/Baby Goats as the other narrative task.

Relevant for the present study are the ten comprehension questions which probe the child's ability to verbalize different types of inferences drawn from the story as shown in the pictures. Note that the MAIN comprehension questions are strictly parallel across stories (in contrast to many other narrative materials in the literature). For each story, three questions (D1, D4, D7) probe understanding of the goals of a main story character in each of the three episodes. Three questions (D2, D5, D8) target a character's internal states, each with a follow-up why-question (D3, D6, D9) that aims to elicit the rationale behind the character's internal states. All questions are open *how* and *why* questions.[7] The final question (D10) requires understanding of the whole plotline. For illustration, the ten comprehension questions for Cat are shown in Table 2. The Cat picture sequence is shown in Figure 1.

Figure 1. Small-scale copy of the Cat picture sequence (Gagarina et al., 2012) (reproduced with permission from the publisher)

7. Following the standard MAIN procedure, the follow-up question is only asked when the preceding question is answered correctly, e.g. D3 is only asked when the answer to D2 is correct. Also, in line with standard MAIN procedure, no forced-choice options were given to the child (e.g. a follow-up question with explicit alternatives) when the child did not answer or answered incorrectly.

Table 2. Overview of the ten comprehension questions, MAIN *Cat*

Question	Example (*Cat*)
D1. Episode 1 Goal	Why does the cat jump/leap forward? (picture 1–2)
D2. Episode 1 IST	How does the cat feel? (picture 3)
D3. Episode 1 IST rationale	Why do you think that the cat is feeling [answer D2]?
D4. Episode 2 Goal	Why does the boy hold the fishing rod in the water? (picture 5)
D5. Episode 2 IST	How does the boy feel? (picture 6)
D6. Episode 2 IST rationale	Why do you think that the boy is feeling [answer D5]?
D7. Episode 3 Goal	Why does the cat grab the fish? (picture 5)
D8. Theory of Mind IST	Imagine that the boy sees the cat. How does the boy feel? (picture 6)
D9. Theory of Mind IST rationale	Why do you think that the boy feels [answer D8]?
D10. Overall plotline question	Will the boy be friends with the cat? Why?

Note. IST = Internal state term.

To get an estimate of the children's vocabulary knowledge, in addition to MAIN, we administered vocabulary production and comprehension tasks in both languages. The Swedish and German versions of the Cross-linguistic Lexical Tasks (LITMUS-CLT, hereafter CLT; Haman, Łuniewska, & Pomiechowska, 2015), CLT-SWE (Ringblom, Håkansson, & Lindgren, 2014) and CLT-DE (Rinker & Gagarina, 2014), were used since they are suitable for children aged 3–6 and were constructed to be of comparable difficulty across languages. (No other comparable vocabulary tasks exist for Swedish and German.) The CLT-SWE and CLT-DE are not translations of each other (nor are they translations of a vocabulary test in another language), but they were constructed in parallel fashion, and the test items for each language version were chosen from a common pool of (word) concepts (cf. Haman et al., 2015). Each CLT language version contains four parts: noun comprehension, verb comprehension, noun production, and verb production. In this study, only CLT noun and verb *production* is reported, as we wanted to assess the effect of expressive vocabulary on narrative comprehension.[8] The CLT production parts are picture naming tasks in which the experimenter asks the child to name an object ('What is this?' for noun production) or action ('What is she

8. Vocabulary comprehension scores were close to ceiling in both languages and did thus not vary to the same extent between individual children. The results for the scores on all four parts of the CLTs in both languages are reported elsewhere (Lindgren & Bohnacker, 2020).

doing?' for verb production). Each child received a vocabulary (production) score (out of maximum 60 points) in each language.[9]

4.3 Procedure

Each child told two MAIN stories per language and answered the comprehension questions for these stories.[10] Each child also did a vocabulary task (CLT) in each language. The two languages were always tested in different sessions with at least three days between the sessions. On average, the time between the sessions was eight days. The order of the languages was counterbalanced for the age groups: half of the children in each age group were tested in Swedish first and the other half in German first. All children were tested in Swedish by the first author, a native speaker of Swedish, and in German by either the second author or a female research assistant, both native speakers of German. The experimenter only spoke to the child in the language of testing and acted as if she did not understand the other language.

The two narrative tasks in each language were carried out within one session as part of a larger test battery. Cat/Dog was always done first, followed by the vocabulary task (CLT), which in turn was followed by Baby Birds/Baby Goats. The child never received the same story in the two languages. The stories told by the child in the respective language were counterbalanced within each age group.[11]

The procedure for all narrative tasks was identical, following the standard procedure for MAIN in the 'telling mode': First the child looked at all six pictures at his/her own pace without them being visible to the experimenter. Next, the child proceeded to tell the story based on his/her interpretation of events in the pictures, while first looking at two, then four, then finally all six pictures (the pictures were still only visible to the child at this point). Finally, the child answered the ten comprehension questions.[12] When the comprehension questions were asked, the

---

9. For more information about the scoring of the children's answers, see Lindgren (2018, pp. 82–84) and Lindgren & Bohnacker (2020). Only correct answers in the target language score a point. Lindgren and Bohnacker (2020) provide an overview of the number of German and Swedish CLT test items that were cognates and found better performance on cognate items. This effect was especially pronounced for German test items that were cognate with Swedish.

10. Results for the children's narrative production are reported elsewhere (Lindgren, 2018: Chapter 7).

11. For an overview of the counterbalancing system used, see Lindgren (2018, p. 59).

12. It is important to point out that, with the exception of the studies by Lindgren (2018) and Bohnacker & Lindgren (in press), previous studies have not used Cat/Dog in the 'telling mode', as these stories were originally created to be used for retelling and model story (Gagarina et al., pictures were visible to both child and experimenter. The entire procedure was audio and video recorded. Standardized questions from the Swedish and German versions of MAIN were used with all children. The child received a sticker after each task, and a diploma after having completed both sessions.

4.4 Scoring and categorization of the comprehension questions

The answers to the comprehension questions were transcribed verbatim. A total of 1,840 comprehension questions were to be asked (46 children x 2 languages x 2 stories x 10 questions). In German, the experimenter forgot to ask four questions (one in Cat/Dog, three in Baby Birds/Baby Goats), corresponding to 0.4% missing data. In Swedish, all questions were asked. However, Swedish data for Cat/Dog was missing for one 5-year-old. This child switched to German during her Swedish narrative and, despite efforts of the experimenter, did not switch back to Swedish, and she also answered the comprehension questions in German. These answers were excluded. The data thus consisted of answers to 1,826 comprehension questions (910 in Swedish and 916 in German). For the four missing German questions, sample mean substitution was used,[13] such that for each missing question, a score identical to the mean for that question for the child's age group was substituted.

Responses to the comprehension questions were scored as 1 (correct) or 0 (incorrect). Thus, for each narrative task, the maximum score on the comprehension questions is 10 points. Each child (with the exception of the above-mentioned 5-year-old) received four different comprehension scores, one for each task.

The first author scored all Swedish answers. The German answers were scored by a native German research assistant and then checked by the first author. Unclear cases were discussed with the second author. The children's answers to the comprehension question were varied and did not always correspond to the typical answers found in the MAIN manual (Gagarina et al., 2012). For this reason, detailed scoring guidelines that included general scoring principles and a large number of authentic answers were developed by the second author and her research group. The principles and decisions of these scoring guidelines were based on multiple rounds of systematic discussion of unclear cases found in more than 9,000 answers of 286

2012), modes in which the child first listens to the story told by the experimenter. In the current study, both Cat/Dog and Baby Birds/Baby Goats were used in the telling mode, and narrative comprehension is analyzed for comprehension questions asked after the child has looked at the pictures and told the story.

13. This follows common practice when there is no more than 2% missing data (Widaman, 2006).

mono- and bilingual children in Sweden. The guidelines were followed closely in the current study and the scoring was thoroughly checked against them.[14]

In addition to scoring the children's answers as correct or incorrect, we analyzed the incorrect answers for specific questions, where accuracy was relatively low. This analysis consisted of grouping the incorrect answers into types with a view to explore general patterns and discover systematicity with respect to age development in the incorrect answers.

5. Results

5.1 Comprehension scores: Differences between age groups, tasks and languages

The mean scores and standard deviations for the comprehension scores of the two narrative tasks (Cat/Dog; Baby Birds/Baby Goats) in each language are shown in Figure 2.

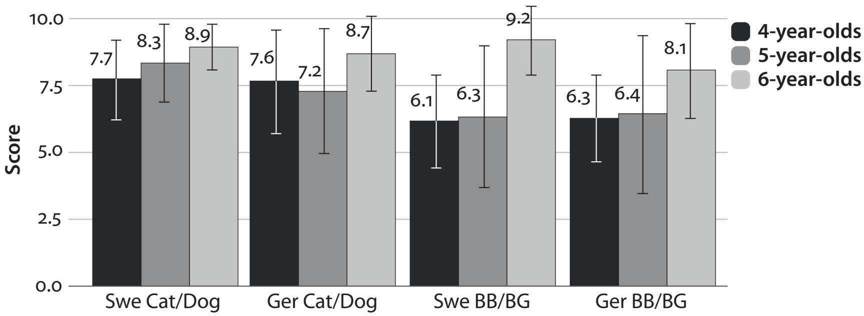

Figure 2. Mean comprehension scores for the narrative tasks in Swedish and German, by age group (Max = 10 points). Error bars show ±1 SD.

There were significant correlations between the two tasks in each language (Swedish: $r = .381, p < .001$, German: $r = .522, p < .001$) and between the same task in the two languages (Cat/Dog: $r = .349, p = .02$, Baby Birds/Baby Goats: $r = .393, p = .006$).

14. We scored our data in the manner described above because MAIN is not norm-referenced and no exhaustive manual for scoring macrostructure and narrative comprehension is available for it as of yet. We believe that our scoring procedure with multiple rounds of discussions, development of guidelines, consensus decisions and subsequent systematic checking of the data is a more valid and solid procedure than commonly reported measures of interrater reliability.

In order to answer the research questions about effects of language, age, and task, a repeated-measures (factorial) ANOVA was carried out with Task (Cat/Dog vs Baby Birds/Baby Goats) and Language as within-child factors, and language of the first testing (Test1) and Age group as between-child factors. In cases of significant interaction effects, we carried out post-hoc analyses (simple effects analyses with Pillai's trace or univariate analyses, depending on the type of interaction effect).

There was a highly significant main effect of Task ($F(1, 39) = 18.288$, $p < .001$, $\eta^2_p = .319$), with higher scores on Cat/Dog than on Baby Birds/Baby Goats. The main effect of Age group was also highly significant ($F(2, 39) = 10.703$, $p < .001$, $\eta^2_p = .354$). Post-hoc tests (Bonferroni-corrected) revealed that the 6-year-olds performed significantly higher than both 4-year-olds and 5-year-olds (ps = .001), but that there was no difference between the two younger groups ($p = 1.00$). The main effects of Language ($F(1, 39) = 2.858$, $p = .099$, $\eta^2_p = .068$) and language of first testing (Test1; $F(1, 39) = 3.073$, $p = .087$, $\eta^2_p = .073$) were not significant. However, the two-way interaction between Language and Test1 was significant ($F(1, 39) = 20.225$, $p < .001$, $\eta^2_p = .341$). This means that the effects of the language of the first testing may be different for the two languages *and/or* that the effect of language may differ depending on the language of the first testing. Simple effects analyses showed that there was no significant effect of test order on scores in Swedish ($F(1, 39) = 1.212$, $p = .278$, $\eta^2_p = .030$), but that the children's scores were significantly higher in German when Swedish was the language of the first testing compared to when German was tested first ($F(1, 39) = 13.294$, $p < .001$, $\eta^2_p = .254$), i.e. for German, there was a training effect. Analyzing the other side of the interaction, we found that when German was the language of the first testing, scores in Swedish were significantly higher than the German scores ($p < .001$, $\eta^2_p = .334$), but when Swedish was the language of the first testing, there was no significant differences between scores in the two languages ($F(1, 39) = 3.853$, $p = .057$, $\eta^2_p = .090$). The interaction is shown in Figure 3.

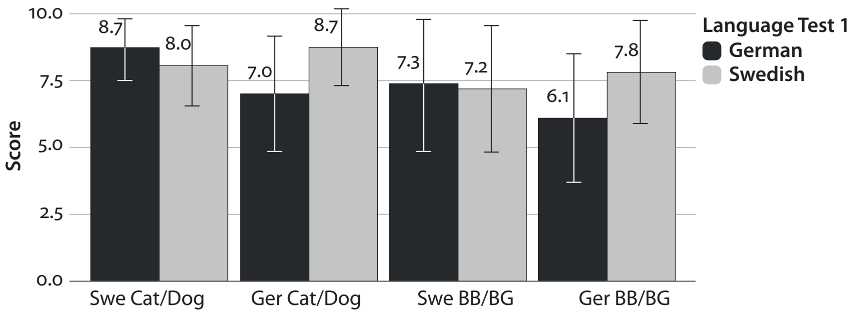

Figure 3. Mean comprehension scores for narrative tasks in Swedish and German, by language of the first testing (Max = 10 points). Error bars show ±1 SD.

The two-way interaction between Task and Group was at the significance level ($F(2, 39) = 3.218$, $p = .05$, $\eta^2_p = .142$). As visual inspection of the data suggests that there might be a different relationship between scores on the two tasks in the different age groups (cf. Figure 2), post-hoc tests were carried out for this interaction effect. Post-hoc tests showed that there were differences between the age groups, with higher scores for the 6-year-olds than the 4- and 5-year-olds for both Cat/Dog ($F(2, 39) = 3.650$, $p = .035$, $\eta^2_p = .158$) and Baby Birds/Baby Goats ($F(2, 39) = 3.218$, $p < .001$, $\eta^2_p = .359$). In line with the results shown in Figure 2, there were significant differences between the narrative tasks, with higher scores on Cat/Dog, for both 4-year-olds ($F(1, 39) = 10.743$, $p = .002$, $\eta^2_p = .216$) and 5-year-olds ($F(1, 39) = 12.845$, $p = .001$, $\eta^2_p = .248$), but not for the 6-year-olds, who performed equally well on both tasks ($F(1, 39) = .205$, $p = .653$, $\eta^2_p = .005$). None of the remaining interaction effects were significant (*p*s > .10).

To summarize, there was a clear effect of age for comprehension of MAIN, with the 6-year-olds outperforming the 4- and 5-year-olds, but with no significant difference between the two younger groups. Test order did not influence the Swedish scores, whereas there was a training effect for German: There was no general significant difference in performance between the languages, but the children performed lower in German than in Swedish when German was tested first. When Swedish was tested first, there was no difference between the languages. Generally, the children reached higher comprehension scores on Cat/Dog than on Baby Birds/Baby Goats; however, there was no difference between these tasks for the 6-year-olds.

5.2 Comprehension scores: Effect of expressive vocabulary knowledge

Next, we investigated the effect of expressive vocabulary knowledge (CLT scores) on the children's narrative comprehension, controlling for the child's age and the language of the first testing.[15] Table 3 gives the summary of the linear regression models for the four different tasks.[16]

15. Note that adding the child's vocabulary comprehension (CLT comprehension score) did not improve model fit, and the effect of vocabulary comprehension was not significant for any of the tasks.

16. Mean scores for the CLT total production scores are reported per age group in Lindgren (2018: Chapter 5) and for verb and noun production separately for all Swedish-German children in Lindgren & Bohnacker (2020).

Table 3. Summary of linear regression models for comprehension scores on the narrative tasks in Swedish and German

	Swe Cat/Dog[a]		Swe BB/BG[b]		Ger Cat/Dog[c]		Ger BB/BG[d]	
	β	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
Intercept	4.90**	1.44	−3.51	2.20	0.83	1.38	−1.11	1.84
Age (months)	0.05*	0.02	0.12***	0.03	0.03	0.02	0.07*	0.03
Vocabulary	0.01	0.03	0.07	0.05	0.10***	0.02	0.07*	0.03
Test1	−0.65	0.37	−0.26	0.56	1.94***	0.42	1.82**	0.55
R^2 (adjusted)	.178		.376		.513		.348	

Note. *** = $p < .001$; ** = $p < .01$; * = $p < .05$.
BB = Baby Birds, BG = Baby Goats, Vocabulary = expressive CLT score in the same language as the narrative task, Test1 = language of the first testing; the models show the effect when the language of the first testing is Swedish.
a. $F(3, 41) = 4.165$, $p = .01$, b. $F(3, 42) = 10.06$, $p < .001$, c. $F(3, 42) = 16.79$, $p < .001$, d. $F(3, 42) = 9.013$, $p < .001$.

For the two tasks in Swedish, age, but not the expressive vocabulary score, significantly predicted the narrative comprehension score. The age effect was somewhat stronger for Baby Birds/Baby Goats as seen by the larger β-value, which may be linked to the level of difficulty of the two tasks, with Cat/Dog being somewhat easier (recall Figure 2). Since the Cat/Dog comprehension scores are higher irrespective of age, there is thus less room for a developmental increase with age. This argument is supported by the results for the same narrative tasks in German; whereas age was not a significant predictor of the German Cat/Dog comprehension scores, it was significant for the comprehension of German Baby Birds/Baby Goats. For both German tasks, the child's German expressive vocabulary score significantly predicted the comprehension scores. The effect of expressive vocabulary is thus not the same in the two languages. Additionally, as already shown above (Section 5.1), the language testing order significantly influenced the German narrative comprehension scores, but not the Swedish ones, with higher scores in German when Swedish was the language of the first testing.

5.3 The performance of individual children

In this section, we take a closer look at individual children, comparing high and low performers in Swedish and German, and report on an older child with a large difference between scores in the two languages. In Table 4, we see a high- and a low-scoring 4-year-old answering the questions about the Dog story in *German*.

Table 4. High- and low-scoring 4-year-olds, MAIN Cat/Dog, German

Question	BiGer4-17 (4;7), Dog	Score	BiGer4-18 (4;5), Dog	Score
D1	*weil der will die Maus fangen will* 'because he wants to catch the mouse'	1	*der hat eine Mause gesehen* 'he has seen a mouse'	0
D2	*traurig* 'sad'	1	*platsch autsch* 'splash ouch'	0
D3	*weil der will die Maus, weil der kann nicht durch den Loch kommen* 'because he wants the mouse, because he cannot come through the hole'	1	–	0
D4	*weil zu den Ballon zu finden, zu nehmen* 'because to find the ball, to take'	1	no answer	0
D5	*fröhlich* 'happy'	1	*der hat einen Ballon* 'he has a balloon'	0
D6	*weil der hat jetzt wieder den Ballon* 'because he now has the balloon back'	1	–	0
D7	*weil der ist hungrig* 'because he is hungry'	1	*der Hund ist hungrig* 'the dog is hungry'	1
D8	*traurig* 'sad'	1	*böse* 'angry'	1
D9	*weil der Hund die Würstchen aufessen* 'because the dog eat the sausages'	1	*die seine Würstchen* 'they his sausages'	0
D10	*nein, weil der hat nämlich die Würstchen aufgegessen* 'no, because he has in fact eaten the sausages'	1	*ich glaube das, ich weiß nicht* 'I think so, I don't know'	0
Total		**10**		**2**

Note. '–' means that the question was not asked, as per the MAIN manual (Gagarina et al., 2012), because the answer to the preceding question was wrong.

These two 4-year-olds are almost the same age, and both attend a Swedish-German bilingual preschool. However, their German answers and the resulting scores are strikingly different. BiGer4-17 provides informative answers and scores high for her age (10 points). Whilst her answers are not always targetlike with regard to form (morphology and syntax), this does not affect the macrostructural score, as the answers are easily comprehensible and match the expected response. By contrast, BiGer4-18 is the child who scores lowest out of all children on MAIN comprehension (2 points, cf. age group mean $M_{4yrs} = 7.6$). He often gives very short answers, and whilst mostly grammatical, they contain far less information (e.g. 'splash ouch', 'I don't know'), and consequently, he receives a very low comprehension score in German. In Swedish, both of these children scored 6 points on the Cat story, which is parallel to the Dog story. BiGer4-17 was tested in Swedish first and BiGer4-18 in German first. This difference in language testing order could have affected the children's scores. However, there are other differences between the children as well. First, concerning expressive vocabulary, while their scores for Swedish vocabulary are comparable and average for their age (BiGer4-17: 46/60, BiGer4-18: 42/60, $M_{\text{CLT-Swe 4yrs}} = 42.1$), in German, BiGer4-17 performed substantially higher (48/60) than BiGer4-18 (28/60) (cf. $M_{\text{CLT-Ger 4yrs}} = 39.1$). There thus seems to be a link between their vocabulary scores in German and their performance concerning story comprehension in German, in line with the results shown in Section 5.2. Although these children both have one parent who is a native speaker of German, and their parents report that they follow the one parent-one language (OPOL) principle when speaking to the child, the children's German language skills are rated quite differently by the parents, BiGer4-17's as 'very good' and BiGer4-18's as 'poor'. The children also differ in terms of the language they speak to their parents: BiGer4-17 follows the OPOL-principle, whereas BiGer4-18 speaks mostly Swedish with both parents. Whether this may be the cause of the difference in language skills between the children or the result thereof cannot be determined here. However, the two children clearly differ in how well they speak German, and this is reflected in their narrative comprehension scores on MAIN.

In Table 5, we see examples of high- and low-scoring 4-year-olds for comprehension of the *Swedish* Cat/Dog.

Table 5. High- and low-scoring 4-year-olds, MAIN Cat/Dog, Swedish

Question	**BiGer4-10 (4;1), Dog**	**Score**	**BiGer4-01 (4;0), Cat**	**Score**
D1	*han vill fånga musen* 'he wants to catch the mouse'	1	*för att han ska jaga fjärilen* 'because he is going to chase the butterfly'	1
D2	*arg* 'angry'	1	*att det kliar* 'that it itches'	0
D3	*han krockar* 'he crashes'	1	–	0
D4	*för att ta ballongen* 'in order to take the balloon'	1	*för att han ska fiska* 'because he is going to fish'	0
D5	*glad* 'happy'	1	*att han har bollen i handen* 'that he has the ball in his hand'	0
D6	*för han har sin ballong* 'because he has his balloon'	1	–	0
D7	*för han är hungrig* 'because he is hungry'	1	*för han gillar fisk* 'because he likes fish'	1
D8	*arg* 'angry'	1	*ledsen* 'sad'	1
D9	*för han vill ta korvarna* 'because he wants to take the sausages'	1	*för katten åt upp alla* 'because the cat ate all'	1
D10	*nej, för han tar alla korvar* 'no, because he takes all the sausages'	1	*nej, för att katten är inte snäll* 'no, because the cat is not nice'	1
Total		**10**		**5**

Note. '–' means that the question was not asked, as per the MAIN manual (Gagarina et al., 2012), because the answer to the preceding question was wrong.

Both these children are among the youngest in the study (4;1 and 4;0, respectively), but whereas BiGer4-10 scores at maximum, with the highest score out of all the 4-year-olds on this task in Swedish (10 points, $M_{\text{Cat/Dog 4yrs}} = 7.7$), BiGer4-01 only reaches half this score (5 points). Recall that the difference in German MAIN comprehension between the two other children discussed earlier (BiGer4-17 and BiGer4-18) could be linked to their German language proficiency and use. However, the same explanations do not hold for the difference in comprehension scores between BiGer4-01 and BiGer4-10 in Swedish. Both these children were tested in Swedish first and they had similar (relatively high) expressive vocabulary scores (BiGer4-01: 49/60; BiGer4-10: 44/60, $M_{\text{CLT-Swe 4yrs}} = 42.1$). Both children scored 9 points on the MAIN Cat/Dog comprehension task in German. Interestingly, it is BiGer4-01 who receives more daily input in Swedish, as she has one native Swedish parent, whereas only German is spoken in the household of BiGer4-10. BiGer4-01 is rated to have 'very good' expressive Swedish skills, whereas BiGer4-10 is only rated as 'good' in Swedish and according to the parents prefers speaking German. It is thus not the case that BiGer4-01's low MAIN score can be linked to a parental estimate of low Swedish proficiency or to low Swedish vocabulary scores. In fact, BiGer4-01's incorrect answers on the Swedish task are typical of a young child, for example in mentioning actions instead of emotions (D5 'How does the boy feel here? [in picture 6]', expected answer 'happy/good', BiGer4-01's answer 'that he has the ball in his hand'), and they do not contain any grammatical errors. It is possible that the difference between the two children is simply due to individual variation in story comprehension, which is relatively large at this age, or that it is related to variation in some other ability, which has not been measured independently in the present study, such as Theory of Mind.

An example of an older child with (very) different comprehension scores in the two languages is BiGer6-18, aged 6;6. In German, this child scores 6 points on the narrative comprehension of Cat and 5 points on Baby Birds (cf. $M_{\text{Cat/Dog 6yrs}} = 8.7$; $M_{\text{BB/BG 6yrs}} = 8.1$), but in Swedish reaches the maximum score of 10 points on both Dog and Baby Goats (cf. $M_{\text{Cat/Dog 6yrs}} = 8.9$; $M_{\text{BB/BG 6yrs}} = 9.2$). Her answers to Baby Goats in Swedish and Baby Birds in German are shown in Table 6.

A number of things are notable in the answers of BiGer6-18. First, her Swedish answers are in line with the expected ones and contain no grammatical mistakes. Second, her German answers contain grammatical mistakes, e.g. in verb inflections (incorrect *jagen* 'chase' instead of *jagt* 'chases'), and in word order (e.g. the child uses VO instead of OV word order, incorrect *fangen die Vögel* 'catch the birds' instead of *die Vögel fangen*). Additionally, she responds to four questions with *ich weiß nicht* 'I don't know'. Since the child is well able to answer the same type of questions in Swedish, as shown by her perfect score in Swedish, it is possible that

Table 6. Answers to the MAIN Baby Birds/Baby Goats Swedish and German comprehension questions, BiGer6-18

Question	Swedish (Baby Goats)	Score	German (Baby Birds)	Score
D1	*för att hjälpa den upp* 'to help it up/ out'	1	*die brauch Essen* 'they need food'	1
D2	*rädd* 'afraid'	1	*Hunger* 'hunger'	1
D3	*för det är för djupt för den* 'because it is too deep for it'	1	*ich weiß nicht* 'I don't know'	0
D4	*för den är hungrig och vill äta den* 'because it is hungry and wants to eat it'	1	*die Katze will fangen die Vögel* 'the cat wants to catch the birds'	1
D5	*rädd och ledsen* 'afraid and sad'	1	*Angst* 'fear'	1
D6	*för fågeln biter på rävens svans, och ser arg ut och vill bita den mer* 'because the bird is biting in the fox's tail, and looks angry, and wants to bite it more'	1	*ein Hund jagen die Katze* 'a dog chase the cat'	1
D7	*för den vill hjälpa den* 'because it wants to help it'	1	*ich weiß nicht* 'I don't know'	0
D8	*glad* 'happy'	1	*ich weiß nicht* 'I don't know'	0
D9	*för dom är i säkerhet* 'because they are safe'	1	–	0
D10	*fågeln, för den jagade iväg räven* 'the bird, because it chased the fox away'	1	*keiner, ich weiß nicht.* 'no-one, I don't know'	0
Total		**10**		**5**

Note. '–' means that the question was not asked, as per the MAIN manual (Gagarina et al., 2012), because the answer to the preceding question was wrong.

she cannot formulate an answer in German. Her expressive vocabulary scores in the two languages are equally dissimilar as her MAIN comprehension scores: she scored 50 points (out of 60) on expressive vocabulary in Swedish, but only half this score (25/60) in German (cf. $M_{\text{CLT-Swe 6yrs}} = 49.0$; $M_{\text{CLT-Ger 6yrs}} = 44.8$). When we compared notes (i.e. our impressions from the sessions in the two languages) on this child, it became clear to us that this child had difficulties speaking German, whereas her Swedish proficiency was at a high level. Whilst her expressive German skills were rated as 'good' by the parents, they also stated that she mainly speaks Swedish with both parents, and that she mainly receives her daily input in Swedish. The parents, one native Swedish and one German native speaker, both speak a mix of Swedish and German with the child. It could thus be the case that this child is not used to speaking German. She was also tested in German first, which probably did not help her perform at her best, as she was then also unfamiliar with the tasks.

5.4 Different comprehension questions

Finally, in order to get a more detailed picture of the children's narrative comprehension, we took a closer look at the accuracy rates for the different comprehension questions and report results from qualitative analyses of the incorrect answers to some specific questions, for Cat/Dog (Section 5.4.1) and Baby Birds/Baby Goats (Section 5.4.2), respectively.

5.4.1 *Cat/Dog*

Figure 4 shows the response accuracies in percentage (i.e. the percentage of the children scoring a point per question) for each of the ten comprehension questions for Cat/Dog in Swedish and German. The pattern in the two languages is strikingly similar.

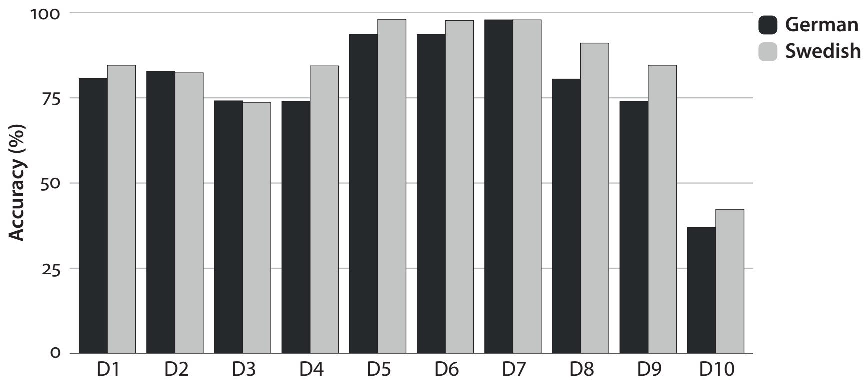

Figure 4. Accuracy (= the percentage of the children answering the question correctly), Cat/Dog comprehension questions, by language, all children (*N* = 46)

Accuracy for nine of the Cat/Dog questions (Table 2, Section 4.2) was high (at least around 75%) and in the case of three questions (D5, D6, D7) even close to 100%. Inferential comprehension of goals and of other internal states of protagonists, including their emotions, was thus generally high. In contrast, accuracy on the final question, D10, was substantially lower (German: 37%, Swedish: 42%). The children's overall performance on Cat/Dog comprehension was thus good, except for the final question. The pattern was identical in the two languages. The Cat/Dog final question (D10), which taps into the understanding of the overall plotline, asks whether the boy would become friends with the cat/dog and includes a follow-up question probing the reasons why/why not. The expected answer is 'no, because the cat/dog ate the boy's fish/sausages' (cf. picture 6 in Figure 1).[17] Why do the children perform so low on this specific question? Is it because they have not understood the plotline, where the cat/dog steals the boy's fish/sausages? The very high response accuracy for the other questions suggests that the children have understood the story. Most children also provided a correct answer to the question (D8/D9) about the hypothetical reaction of the boy if he were to turn around and see the cat/dog (eating the fish/sausage). They thus understood that the boy would be angry at the cat/dog. Still, they were not able to make the link between the animal's behaviour and the potential for friendship with the boy. So, what did they answer instead of the expected answer on D10, and is there any development with age?

We carried out a qualitative analysis of the incorrect responses and experimented with different ways of grouping them. Finally, we assigned the incorrect responses to one of the following three answer types: (i) *I don't know* or no explanation given, (ii) reasonable (logical) explanations, (iii) cryptic or nonsensical explanations (see below). The types of answers in Cat/Dog D10 are shown in Table 7 for Swedish and German combined.[18] As D10 is asked in connection with the last picture in Cat and Dog, this picture (Picture 6) is shown in Figure 5.

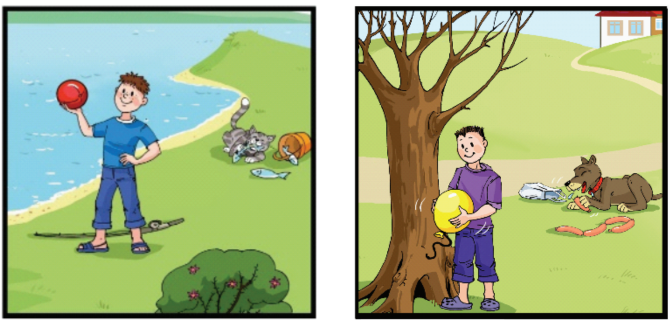

Figure 5. Small-scale copies of picture 6 in MAIN *Cat* (left) and *Dog* (right)

17. Answers of the type 'yes, because the boy will forgive the cat/dog' are accepted as well.

18. Patterns were very similar in the two languages and therefore the answers were combined. The total thus consists of 2 answers per child, one in each language (e.g. for the 4-year-olds, 2 answers x 14 children gives a total of 28). Note that one 5-year-old did not answer the questions for the Swedish task, and hence, the total here is 31 instead of 32.

Table 7. Types of answers Cat/Dog D10, by age group

Answer type	4-year-olds		5-year-olds		6-year-olds	
	%	N	%	N	%	N
Correct answer	32.1	9	38.7	12	46.9	15
Incorrect answer	67.9	19	61.3	19	53.1	17
– ‘*I don’t know*’/no explanation	35.7	10	29.0	9	18.8	6
– Reasonable explanation	14.3	4	12.9	4	31.3	10
– Cryptic/nonsensical explanation	17.9	5	19.4	6	3.1	1
Total	**100.0**	**28**	**100.0**	**31**	**100.0**	**32**

Note. N = number of answers (German and Swedish combined).
Question D10, Cat/Dog: *Will the boy be friends with the cat/dog? Why?*
Expected answer: *No, because the cat/dog ate the boy’s fish/sausages.*

As seen in Table 7, there is some age development in the percentage correct answers to Cat/Dog D10, but not a very large one: the 6-year-olds answer the question correctly to a somewhat higher degree than the 4-year-olds, with the 5-year-olds in the middle. More notable are the differences between the age groups in the *types* of incorrect answers given.

While the 4- and 5-year-olds relatively often gave no explanation for their answer or gave cryptic/nonsensical explanations, this was much less common in the 6-year-olds. Such cryptic or nonsensical explanations include e.g. *nej, därför det här såg ut som en hundvalp* ‘no [they won’t be friends] because it looked like a puppy’ [Swe.], or *ja, weil das ist eine böse Katze* ‘yes [they will be friends], because it is an angry cat’ [Ger.], or *doch, ich glaube, die möchte diese Fisch* ‘yes [they will be friends, because] I think she (the cat) wants the fish’ [Ger.]. By contrast, the 6-year-olds’ incorrect answers were mostly reasonable, logical explanations. Incorrect, but logical explanations are for example answering that the boy will be friends with the cat/dog because it is in fact the boy’s cat or dog (e.g. *det kanske var hans katt* ‘it was maybe his cat’ [Swe.], *vielleicht gehört dem Mann der Hund* ‘perhaps the dog belongs to the man’ [Ger.]). Whilst such answers are not based on the events depicted in the Cat/Dog story, they are based on (accurate) world-knowledge, where dogs do tend to belong to people, and if a dog is out walking close to a person, it is likely that it is that person’s dog, and if you have a cat/dog as a pet, you tend to be friends with your pet, etc. Such explanations are thus to a larger extent ‘correct’, even though they are not the expected answers in the context of this narrative task. On this particular question, the answers of the 6-year-olds thus give evidence of increased inferential understanding and reasoning, even though this does not show in the purely quantitative results. Next, we move on to the answers for Baby Birds/Baby Goats.

5.4.2 *Baby Birds/Baby Goats*

Figure 6 shows the response accuracies for the Baby Birds/Baby Goats comprehension questions in the two languages. Questions targeting the inferential comprehension of protagonists' goals (D1, D4, D7) were generally understood well (>75%), just as was the case for Cat/Dog. Other internal states of protagonists, including their emotional reactions, presented a more mixed picture and they were generally less well understood in Baby Birds/Baby Goats than in Cat/Dog. Overall performance, too, is lower for Baby Birds/Baby Goats compared to Cat/Dog, with the clear exception of D10, where accuracy in Baby Birds/Baby Goats is nearly three times higher than in Cat/Dog (D4 has higher accuracy in Baby Birds/Baby Goats too). For three questions (D3, D8, D9), performance on Baby Birds/Baby Goats is substantially lower, with only around 50% or even fewer children answering these questions correctly. The pattern is very similar in German and Swedish.

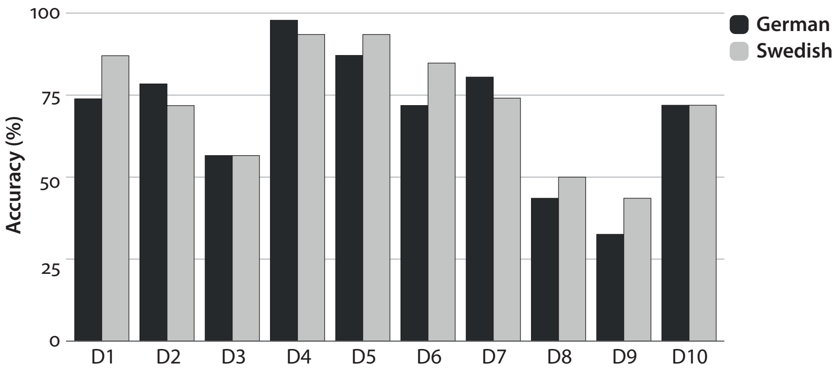

Figure 6. Accuracy (= the percentage of the children answering the question correctly), Baby Birds/Baby Goats comprehension questions, by language, all children ($N = 46$)

Let us now consider the answers to these low accuracy questions (D3, D8, D9) in turn. As D3 is the follow-up question to D2 (both are asked in connection with picture 1), and is only asked if D2 is answered correctly, we first need to look at answers to D2. Figure 7 shows picture 1 from the two stories.

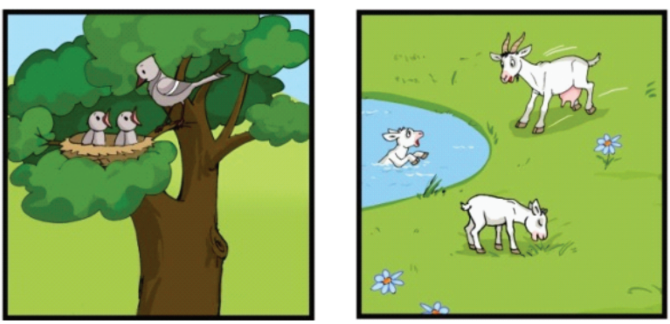

Figure 7. Small-scale copies of picture 1 in MAIN *Baby Birds* (left) and *Baby Goats* (right)

Correct answers to D2 (*How do the baby birds feel here/How does the baby goat feel here* [in picture 1]?) are 'bad', 'not good', 'hungry' (for Baby Birds), and 'bad', 'not good', 'scared' (for Baby Goats). The answers of our participants are broken down in Table 8 and show a clear development with age, with almost all 6-year-olds answering D2 correctly, whereas only 59%–71% of the younger children do so (note however that the 4-year-olds have a somewhat higher accuracy than the 5-year-olds). A qualitative analysis shows that the wrong answers are a mix of 'I don't know', clearly positive emotions (*glada* 'happy' [Swe.], *gut* 'good' [Ger.]), other emotions (i.e. emotions that are not positive, but not correct either, e.g. *ein bisschen müde* 'a bit tired' [Ger.], *dom känner sig lite löjliga* 'they feel a bit silly' [Swe.]), and descriptions of events (e.g. *dom säger hej* 'they say hi' [Swe.], *die kann nicht schwimmen* 'she can't swim' [Ger.]). In addition to showing a higher accuracy than the 4- and 5-year-olds, the 6-year-olds never give positive emotions as answers to D2, whereas the younger children do.

Table 8. Types of answers, Baby Birds/Baby Goats D2

Answer type	**4-year-olds**		**5-year-olds**		**6-year-olds**	
	%	N	%	N	%	N
Correct answer	71.4	20	59.4	19	93.8	30
Incorrect answer	28.6	8	40.6	13	6.3	2
– '*I don't know*'	3.6	1	6.3	2	6.3	2
– Clearly positive emotion (*good, happy*)	7.1	2	15.6	5	0.0	0
– Other (negative) emotion (*wet, silly, tired*)	7.1	2	6.3	2	0.0	0
– Action/description of events (*they say hi, she can't swim*)	10.7	3	12.5	4	0.0	0
Total	**100.0**	**28**	**100.0**	**32**	**100.0**	**32**

Note. N = number of answers (German and Swedish combined).
Question D2, Baby Birds/Baby Goats: *How do the baby birds/does the baby goat feel here* [in picture 1]*?*
Expected answer: *bad, not good, hungry* (for Baby Birds); *bad, not good, scared* (for Baby Goats).

Having seen an age development in the answers to D2, let us now move on to the follow-up question D3 (*Why do you think that the baby birds are feeling/that the baby goat is feeling* [Answer D2]*?).* Table 9 shows the types of answers given by the children on D3. The expected answers include '(bad) because they are hungry' (Baby Birds), '(scared/afraid) because s/he cannot swim' (Baby Goats), '(bad/scared) because s/he has fallen into the water' (Baby Goats). Here, there is also a clear age development; the 6-year-olds more often answer the question correctly (75%) than the 4- and 5-year-olds (43%–50%). We tried to classify the incorrect responses to D3 into subtypes but they were of rather mixed types. Most commonly, children gave no answer or said 'I don't know'. We could not discern any clear pattern with age.

Table 9. Types of answers, Baby Birds/Baby Goats D3

Answer type	4-year-olds		5-year-olds		6-year-olds	
	%	N	%	N	%	N
D2 incorrect (D3 not asked)[19]	28.6	8	40.6	13	6.3	2
Correct answer	42.9	12	50.0	16	75.0	24
Incorrect answer	28.6	8	9.4	3	18.8	6
– No answer/'*I don't know*'	14.3	4	3.1	1	9.4	3
– Because of cat/fox	3.6	1	6.3	2	0.0	0
– Other answer	10.7	3	0.0	0	9.4	3
Total	**100.0**	**28**	**100.0**	**32**	**100.0**	**32**

Note. N = number of answers (German and Swedish combined).
Question D3, Baby Birds/Baby Goats: *Why do you think that the baby birds are feeling/that the baby goat is feeling* [Answer D2]*?*
Expected answer: *because they are hungry/haven't got food* (for Baby Birds); *because s/he has fallen into the water and can't swim* (for Baby Goats).

Finally, we investigated the answers to the Baby Birds/Baby Goats comprehension question D8 (*Imagine that the dog/bird sees the birds/goats. How would the dog/bird feel?*), asked in connection with the final picture. This picture (Picture 6) is shown in Figure 8.

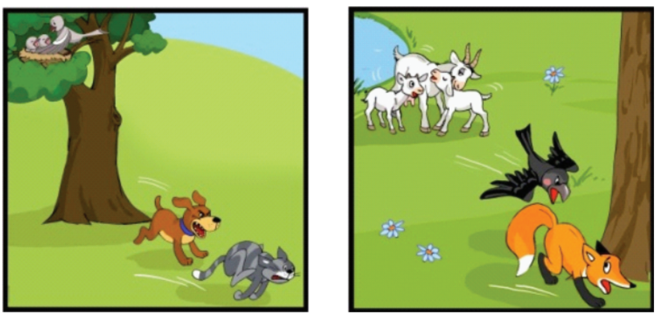

Figure 8. Small-scale copies of picture 6 in MAIN *Baby Birds* (left) and *Baby Goats* (right)

Correct answers to question D8 are those expressing a positive emotion, such as *gut* 'good' [Ger.], *glücklich* 'happy' [Ger.], *snäll* 'nice' [Swe.], or *stolt* 'proud' [Swe.]. The answer types are broken down in Table 10.[20] Here, we see a clear age development,

19. Following standard MAIN procedure, D3 was only asked when then answer on D2 was correct.

20. According to standard MAIN procedure, D9 is only asked when D8 is answered correctly. Of those children who did answer D8 correctly (which the majority of the children did not do, Figure 5), very few children (3 in Swedish, 4 in German) answered D9 incorrectly. The incorrect answers were 'because he wants to eat them' (3), 'because he chases the fox' (2), 'I don't know' (1), and 'because maybe he is friends with the goats' (1).

where the 6-year-olds answer this question correctly more often than the 4- and 5-year-olds. We qualitatively analyzed the incorrect answers on D8 and grouped them into five categories, inspired by Bohnacker (2016) and Bohnacker and Lindgren (in press).

Table 10. Types of answers, Baby Birds/Baby Goats D8

Answer type	4-year-olds		5-year-olds		6-year-olds	
	%	N	%	N	%	N
Correct answer	32.1	9	37.5	12	68.8	22
Incorrect answer	67.9	19	62.5	20	31.3	10
– Clearly negative emotion *(angry, sad, not good)*	46.4	13	37.5	12	12.5	4
– Other emotion *(a bit crazy, slow)*	0.0	0	6.3	2	0.0	0
– Physical state *(hungry/wanting to eat)*	10.7	3	12.5	4	6.3	2
– Action *(going to eat them, is saving the birds)*	3.6	1	3.1	1	0.0	0
– *'I don't know'*	7.1	2	3.1	1	12.5	4
Total	**100.0**	**28**	**100.0**	**32**	**100.0**	**32**

Note. N = number of answers (German and Swedish combined).
Question D8, Baby Birds/Baby Goats: *Imagine that the dog/bird sees the birds/goats. How would the dog/bird feel?* Expected answer: *good, proud, like a hero*.

Most of the incorrect answers on Baby Birds/Baby Goats D8 were clearly negative emotions, e.g. *böse/arg* 'angry' [Ger./Swe.], *inte glad* 'not happy' [Swe.] or *nicht gut* 'not good' [Ger.]. Such answers show that the child has not completely understood the role of the dog/crow as the helper or 'saviour' in the story, but instead answers according to the facial expression of the dog/crow in the last picture (an angry face, as s/he is chasing the cat/fox away, see Figure 8). Similarly, answers in the form of a physical state, e.g. the dog/crow being *hungrig* 'hungry' [Swe./Ger.] or wanting to eat the birds/goats (*hunden vill äta upp dom* 'the dog wants to eat them (the birds)' [Swe.]), show a misunderstanding of the role of the dog/crow.

6. Discussion and conclusion

This paper has investigated aspects of story comprehension in German-Swedish bilingual 4- to 6-year-olds in both their languages. The children answered comprehension questions for two picture-based stories (MAIN Cat/Dog and Baby Birds/Baby Goats) in each language. The research questions concerned development with age (RQ1), differences between the languages both with respect to age development (RQ1) and in scores (RQ2), task effects (differences in performance between Cat/Dog and Baby Birds/Baby Goats) (RQ3), and effects of general expressive vocabulary on story comprehension when controlling for age and order of testing (RQ4). These first four questions all concerned overall narrative comprehension scores. Finally, we analyzed responses to individual comprehension questions in depth, focusing on age development in the types of answers given by the children (RQ5). Our results for the five research questions will now be summarized and discussed in turn.

First, we could document a clear effect of age in narrative comprehension (overall scores). The 6-year-olds performed significantly better than the 4- and 5-year-olds, with no significant difference between the two younger groups. The development with age was similar in both languages and was consistent across tasks. The effect of age is in line with earlier studies, showing a clear development between age 4–5 and 6–7 (e.g. Bohnacker, 2016; Bohnacker & Lindgren, in press). There are thus indications that an important step in narrative development is taken in the late preschool and early school years. At age 6, comprehension of the Cat/Dog and Baby Birds/Baby Goats narratives is at a relatively high level, with goals being nearly always accurately inferred on both narrative tasks. The results of the present study thus support the conclusion by Bohnacker and Lindgren (in press) that children's goal comprehension in MAIN should be well-developed by age 6.

Second, the main effect of language was not significant, but the children performed lower in German than in Swedish when German was tested first. Such an effect of test order was only found for German. When Swedish was tested first, no difference was found between the languages. The lack of a general difference in scores between the languages is in line with previous studies investigating narrative comprehension using MAIN (Bohnacker, 2016; Kapalková et al., 2016; Roch et al., 2016). Effects of order of testing on narrative comprehension have rarely been investigated. Bohnacker (2016), who studied Swedish-English bilinguals, did not find any significant effect of the order of testing. However, she did not investigate the interaction between language and testing order statistically, which may be the reason for the difference in results. A possible explanation for why a training effect was only found for German in the present study lies in the fact that, for most children, according to parental report, German is the language to which they are exposed less and which is rated by the parents to be the weaker one (see Section 4.1). Carrying out these narrative tasks in German, which entailed having to speak only German to a stranger, may thus have been more demanding for the children than doing the same in Swedish, due to (somewhat) lower language proficiency in German. Having already told narratives and answered questions in Swedish may have made the German session easier, since the children then knew what to expect from the tasks. In short, having been tested in Swedish first may have helped the children to show their full capability in the German session, leading to equal performance in both languages.

Third, the children generally performed better on Cat/Dog than on Baby Birds/Baby Goats, although there was no significant difference between the tasks for the 6-year-olds. The reason why the 6-year-olds performed similarly on both tasks most likely lies in their overall high performance, with mean scores that are approaching ceiling. The better performance on Cat/Dog comprehension compared with Baby Birds/Baby Goats is in line with results for Swedish monolinguals reported in Bohnacker and Lindgren (in press). It should be noted that the present study reported results for 10 comprehension questions, whereas Bohnacker and Lindgren only investigated the first nine MAIN comprehension questions. In the present study, the performance on the tenth and final question (D10), was lower for Cat/Dog than for Baby Birds/Baby Goats (Figures 4 and 6), and notably lower than on all the other Cat/Dog questions (Figure 4). Including D10 in Cat/Dog thus lowered the children's overall performance. In contrast, accuracy on D10 in Baby Birds/Baby Goats was higher; generally, the children performed well on this question (Figure 6). Including this question in Baby Birds/Baby Goats thus had less of an impact on the overall comprehension scores. Despite this, the German-Swedish children scored significantly higher on Cat/Dog. The results of the present study thus support the conclusion drawn by Bohnacker and Lindgren (in press) and Bohnacker et al. (this volume), namely that comprehension of Cat/Dog is in fact easier than Baby Birds/Baby Goats, when both sets of questions are asked after the child has told the story. This suggests that differences found between the tasks when Cat/Dog was used in the *retelling* mode (i.e. when the child first listened to the story and then retold it) and when Baby Birds/Baby Goats was used in the story telling mode (e.g. Maviş et al., 2016; Roch et al., 2016) may not only have been due to a story modality task effect (retelling vs telling), but also the result of the comprehension questions of Cat/Dog being easier. The finding that comprehension of Cat/Dog is easier than Baby Birds/Baby Goats is noteworthy, especially in view of the fact that the MAIN tasks, including the questions, were designed to be comparable (see Gagarina et al., 2015, p. 256). Future studies that plan to investigate differences in comprehension between MAIN retelling and MAIN story telling should therefore not compare Cat/Dog with Baby Birds/Baby Goats, but use the same narrative task (e.g. Cat/Dog *or* Baby Birds/Baby Goats) in both modes.

Fourth, in German, but not in Swedish, CLT expressive vocabulary scores significantly predicted narrative comprehension scores. The effect of expressive vocabulary was thus not the same in the two languages. For Swedish, there were instead clear and significant effects of age for both narrative comprehension tasks. The age effect was somewhat stronger for Baby Birds/Baby Goats. Whereas age was not a significant predictor of the German Cat/Dog comprehension scores, it was significant for the comprehension of German Baby Birds/Baby Goats. The reason why age is a predictor for the scores on German Baby Birds/Baby Goats may be related to the level of difficulty of the two tasks. Since the Cat/Dog comprehension scores are higher irrespective of age, there is less room for an increase with age. One possible reason why expressive vocabulary was a significant predictor in German but not in Swedish is that, in Swedish, these children's expressive vocabulary scores are significantly predicted by the child's age, but this is not the case for German.[21] In the majority language Swedish, the child's age is thus the central factor explaining the child's language proficiency, whereas in the minority language German, there is not the same clear link between age and language skills. As a group, the children in the present study perform better on expressive vocabulary in Swedish than in German (Lindgren & Bohnacker, 2019). Generally, both in terms of our impressionistic judgment of the children's level of linguistic complexity and the grammatical errors in their narratives, and according to parental estimates (see Section 4.1), most children are (at least somewhat) more proficient in Swedish than in German. This may mean that most children's language skills in Swedish were above the threshold needed for being able to answer the story comprehension questions in a comprehensible manner, but that this was not equally the case in German. Above this threshold, whether a child has a slightly lower or slightly higher expressive vocabulary score does not matter for the child's ability to answer and score on the MAIN comprehension questions. What matters when the child has reached this threshold in his/her language skills is the child's ability to understand the story, which is related to cognitive maturity and age. However, if the child's language skills are too rudimentary, s/he may not be able to convey the understanding of the picture-based story to the listener in the form of (correct) answers to probe questions, as such answers must be expressed verbally. The documented effect of expressive vocabulary on narrative comprehension scores indicates that the vocabulary tests pick up on this general language proficiency, which is necessary to verbally demonstrate narrative comprehension. If there were no links between the comprehension scores and the vocabulary knowledge, one would expect a bilingual child's narrative comprehension scores to be (close to) identical in both languages, irrespective of proficiency in the respective languages. Both our quantitative analysis and the discussion of individual children (Sections 5.1–5.3) show that this is not the case.

21. The relationship between CLT expressive vocabulary and age in the same German-Swedish bilinguals has been investigated in Lindgren & Bohnacker (2020); here, the child's age was a significant predictor for Swedish expressive vocabulary scores but not for German scores.

Finally, we took a closer look at the children's performance on different comprehension questions. Patterns in the two languages were strikingly similar. For Cat/Dog, performance was high on all questions except the final one, D10. There was some age development in accuracy on this question, but the main difference between the 6-year-olds and the younger children was found in the types of incorrect answers. The 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. For Baby Birds/Baby Goats, overall performance was lower than for Cat/Dog (with the notable exception of D10). Accuracies on the question pairs D2/D3 and D8/D9 were especially low. With the exception of D2, where positive emotions were never given as answers by 6-year-olds, no clear development with age concerning the *types* of incorrect answers was found, but there was a clear *age* development in accuracy; the 6-year-olds mostly answered these questions correctly, which was not the case for the younger children. Response patterns for the Baby Birds/Baby Goats question D8 which queried the emotional reaction of a protagonist at the end of the story plot are comparable to those reported in Bohnacker and Lindgren (in press) for Swedish monolinguals and English-Swedish bilinguals. Here, just as in the present study, 4- and 5-year-olds mostly answered D8 with a negative emotion, when the expected answer is a positive one. Such answers show that the children did not take the whole plotline into account, but instead focused on the character's facial expression in the last picture (e.g. the dog looking angry) and inferred the character's emotional state from this expression (e.g. being angry). The 6-year-olds drew upon the whole of the plotline to a larger extent and could thus more often answer the question correctly. Even though our German-Swedish bilinguals performed somewhat better than the English-Swedish bilinguals and monolinguals in Bohnacker and Lindgren (in press), a substantial number of children at age 6 still struggled to infer the emotional reaction of the protagonist on D8.

We therefore fully agree with Bohnacker and Lindgren's (in press) conclusion that not every aspect of inferential understanding probed on MAIN can be expected to be mastered by age 6. The age/performance benchmarks suggested by Bohnacker and Lindgren (in press) for MAIN comprehension seem to hold for our German-Swedish bilinguals as well. When findings are the same across age-matched children of different language combinations, this suggests that they generalize. Note however that both the German-Swedish participants of the current study and the English-Swedish and monolingual Swedish children reported in Bohnacker and Lindgren (in press) have very similar backgrounds: they come from high-SES backgrounds (measured via parental education levels), they have all attended Swedish preschool from an early age, and they are regularly exposed to storytelling and book-reading activities in the home and/or at preschool. It remains to be seen whether age-matched children from very different backgrounds would perform comparably on MAIN narrative comprehension.[22]

What do these results then tell us about children's inferential comprehension of picture-based stories? They tell us that 4- to 6-year-olds are well able to infer (and express) the goals and emotions of story characters from purely visually presented stories when prompted, and they are able to do so even though they have not listened to the story before. The method employed here, where carefully designed picture sequences are visible to the child throughout the procedure, no doubt unburdens the child from some of the memory demands imposed by methods used in many other studies, where the child first listens to the story with or without pictorial support and is then asked to recall/retell the story and/or answer comprehension questions *without the pictures present* (e.g. Bishop & Adams, 1992; Lepola et al., 2012; Letts & Leinonen, 2001; Lynch et al., 2008; Omanson et al., 1978; Paris & Paris, 2003; Stein & Glenn, 1979; Trabasso et al., 1984; van den Broek et al., 2005; and the studies in the meta-analysis by Filiatrault-Veilleux et al., 2015). As pointed out in Section 1, understanding a story is not so much about recall of factual details (presented orally or visually), but about inferential comprehension of the relations between story events, including the reasons for a character's actions, understanding the plot of the story, and interpreting it against a backdrop of prior knowledge – knowledge of facts of the world, familiarity with social interaction and cause-and-effect relationships, and knowledge of narrative schemata (Burris & Brown, 2014; Stein & Glenn, 1979; Trabasso & Rodkin, 1994; van den Broek et al., 2005). Whilst the children in this sense generally showed relatively good overall understanding of the picture-based stories in the present study, we also found a clear development with age, and it was not the case that the children had mastered every aspect of inferential narrative comprehension at age 6. Note that very similar types of (adultlike and non-adultlike) inferences were made across languages and at very similar frequencies; thus, children's inferential comprehension was more dependent on age (i.e. general cognitive development) than on language. This is only to be expected, since inferencing skills – even when assessed via verbalized responses – are part of general cognition, and inferencing skills should therefore manifest similarly and/or transfer between languages. However, it must also be said that children need to have a certain minimum proficiency level in a language to be able to verbalise and thus communicate their inferential comprehension of a story

22. Interestingly, Bohnacker et al. (this volume) found similar results for both languages of 100 Turkish-Swedish bilinguals, who came from lower SES-backgrounds and had received less exposure to Swedish compared to the German-Swedish bilinguals of the present study.

to the listener/experimenter, as it is not possible to explain how characters feel and think, and why they feel or think that way, by purely nonverbal means (such as pointing to a picture or drawing a happy/sad face).

To conclude, in this study we have shown that there is a clear development from age 4 to age 6 in German-Swedish bilinguals' narrative comprehension *in both languages*. Importantly though, the children's performance depends on both the narrative task used, with higher scores for MAIN Cat/Dog than Baby Birds/ Baby Goats, and on the specific aspect of story structure probed, as response accuracy was found to vary substantially between different comprehension questions. Response patterns to individual questions were strikingly similar in both languages, suggesting that they may generalize across languages. Expressive vocabulary may also influence children's narrative comprehension: Here, the effect of vocabulary knowledge on narrative comprehension as probed on MAIN may depend on the children's overall level of language skill as well as the relationship between the children's vocabulary and their age.

Acknowledgements

The authors would like to thank Natalia Gagarina and three *SiBiL* reviewers for their helpful suggestions and criticisms, Valerie Reichardt for assistance with data collection and transcription, and the children, parents, preschools and schools for participating in the study.

Funding

Funded by: Swedish Research Council.
Award ID: VR2013-1309.
Award recipient: Ute Bohnacker.
Statement: This work was partially supported by a grant from the Swedish Research Council (VR2013-1309) to Ute Bohnacker.

References

Astington, J. W., & Pelletier, J. (2005). Theory of mind, language, and learning in the early years: Developmental origins of school readiness. In B. D. Homer & C. S. Tamis-LeMonda (Eds.), *The development of social cognition and communication* (pp. 312–352). Mahwah, NJ: Lawrence Erlbaum Associates.

Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. Hove: Psychology Press.

Bishop, D. V. M., & Adams, C. (1992). Comprehension problems in children with specific language impairment: Literal and inferential meaning. *Journal of Speech and Hearing Research*, 35(1), 119–129. https://doi.org/10.1044/jshr.3501.119

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

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Bohnacker, U., & Lindgren, J. (in press). MAIN story comprehension: What can we expect of a typically developing child? In S. Armon-Lotem & K. K. Grohmann (Eds.), *LITMUS in action: Cross-comparison studies across Europe*. Amsterdam: John Benjamins. Preprint retrieved from <http://uu.diva-portal.org/smash/get/diva2:1348899/FULLTEXT01.pdf> (18 June, 2019).

Bohnacker, U., Lindgren, J. & Öztekin, B. (this volume). Bilingual Turkish-Swedish children's understanding of MAIN picture sequences: Individual variation, age, language and task effects. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins.

Boudreau, D. M. (2007). Narrative abilities in children with language impairments. In R. Paul (ed.), *Language disorders from a developmental perspective: Essays in honour of Robin S. Chapman* (pp. 331–356). Mahwah, NJ: Lawrence Erlbaum Associates.

Burris, S. E., & Brown, D. D. (2014). When all children comprehend: Increasing the external validity of narrative comprehension development research. *Frontiers in Psychology*, 5 (Article 168). https://doi.org/10.3389/fpsyg.2014.00168

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: An investigation of their narratives and memory. *International Journal of Language and Communication Disorders* 43(2), 201–218. https://doi.org/10.1080/13682820701366147

Filiatrault-Veilleux, P., Bouchard, C., Trudeau, N., & Desmarais, C. (2015). Inferential comprehension of 3–6 year olds within the context of story grammar: A scoping review. *International Journal of Language & Communication Disorders*, 50(6), 737–749. https://doi.org/10.1111/1460-6984.12175

Gagarina, N. (2016). Narratives of Russian–German preschool and primary school bilinguals: Rasskaz and Erzählung. *Applied Psycholinguistics*, 37(1), 91–122. https://doi.org/10.1017/S0142716415000430

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics, 56*. Berlin: Zentrum für Allgemeine Sprachwissenschaft.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children disentangling bilingualism from language impairment* (pp. 243–269). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-011

Gibson, T. A., Peña, E. D., & Bedore, L. M. (2018). The receptive–expressive gap in English narratives of Spanish–English bilingual children with and without language impairment. *Journal of Speech Language and Hearing Research*, 61(6), 1381–1392. https://doi.org/10.1044/2018_JSLHR-L-16-0432

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, 13(2), 175–197. https://doi.org/10.1016/S0898-5898(01)00061-4

Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing Cross-Linguistic Lexical Tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment* (pp. 196–239). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-010

Hayward, D., Schneider, P., & Gillam, R. B. (2009). Age and task-related effects on young children's understanding of a complex picture story. *The Alberta Journal of Educational Research*, 54(1), 54–72.

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, 15(1), 58–74. https://doi.org/10.1017/S1366728911000538

Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak–English children: A cross-language comparison. *Applied Psycholinguistics*, 37(1), 145–164. https://doi.org/10.1017/S0142716415000454

Kintsch, W. (1988). The role of knowledge in discourse comprehension. A constructive-integration model. *Psychological Review*, 95, 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, 37(1), 123–144. https://doi.org/10.1017/S0142716415000442

Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4–6-year-old children. *Reading Research Quarterly* 47(3), 259–282. https://doi.org/10.1002/rrq.020

Letts, C., & Leinonen, E. (2001). Comprehension of inferential meaning in language-impaired and language normal children. *International Journal of Language & Communication Disorders*, 36(3), 307–328. https://doi.org/10.1080/13682820110045829

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813313

Liles, B. Z. (1993). Narrative discourse in children with language disorder and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research*, 36(5), 868–882. https://doi.org/10.1044/jshr.3605.868

Lindgren, J. (2018). *Developing narrative competence: Swedish, German-Swedish and Turkish-Swedish children aged 4–6* (Studia Linguistica Upsaliensia 19). Uppsala: Acta Universitatis Upsaliensis.

Lindgren, J. (2019). Comprehension and production of narrative macrostructure in Swedish: A longitudinal study from age 4 to 7. *First Language*, 39(4), 412–432. https://doi.org/10.1177/0142723719844089

Lindgren, J., & Bohnacker, U. (2020). Vocabulary development in closely-related languages: Age, word type and cognate facilitation effects in bilingual German-Swedish preschool children. *Linguistic Approaches to Bilingualism*. 10(5), 587–622. https://doi.org/10.1075/lab.18041.lin

Lynch, J. S., van den Broek, P., Kremer, K- E., Kendeou, P., White, M. J. & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology*, 29, 327–365. https://doi.org/10.1080/02702710802165416

Mäkinen, L. (2014). Narrative language in typically developing children, children with specific language impairment and children with autism spectrum disorder (Unpublished doctoral dissertation). University of Oulu.

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, 37(1), 69–89. https://doi.org/10.1017/S0142716415000429

Omanson, R. C., Warren, W. H., & Trabasso, T. (1978). Goals, inferential comprehension, and recall of stories by children. *Discourse Processes*, 1(4), 337–354. https://doi.org/10.1080/01638537809544444

Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*. Published online 2 February 2018. https://doi.org/10.1080/13670050.2018.1434124

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, 38(1), 36–76. https://doi.org/10.1598/RRQ.38.1.3

Pelletier, J., & Astington, J. W. (2004). Action, consciousness and theory of mind: Children's ability to coordinate story characters' actions and thoughts. *Early Education and Development* 15(19), 5–22. https://doi.org/10.1207/s15566935eed1501_1

Ringblom, N., Håkansson, G., & Lindgren, J. (2014). Cross-linguistic lexical task: Swedish version (CLT-SWE) (Unpublished material).

Rinker, T., & Gagarina, N. (2014). Cross-linguistic lexical task: German version (CLT-DE) (Unpublished material).

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37(1), 49–67. https://doi.org/10.1017/S0142716415000417

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives* (pp. 53–120). Norwood, NJ: Ablex.

Tompkins, V., Guo, Y., & Justice, L. M. (2013). Inference generation, story comprehension, and language skills in the preschool years. *Reading and Writing*, 26(3), 403–429. https://doi.org/10.1007/s11145-012-9374-7

Trabasso, T., & Rodkin, P. C. (1994). Knowledge of goal/plans: A conceptual basis for narrating "Frog where are you?" In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: A crosslinguistic developmental study* (pp. 85–106). Hillsdale, NJ: Lawrence Erlbaum Associates.

Trabasso, T., Secco, T., & van den Broek, P. W. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Lawrence Erlbaum Associates.

Trabasso, T., Stein, N. L., Rodkin, P. C., Munger, M. P., & Baughn, C. R. (1992). Knowledge of goals and plans in the on-line narration of events. *Cognitive Development*, 7(2), 133–170. https://doi.org/10.1016/0885-2014(92)90009-G

Van den Broek, P., Kendeou, P., Kremer, K., Lynch, J. S., Butler, J., White, M. J., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. Stahl & S. Paris (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum Associates.

Westerveld, M. F., Gillon, G. T., & Boyd, L. (2012). Evaluating the clinical utility of the Profile of Oral Narrative Ability for 4-year-old children. *International Journal of Speech-Language Pathology* 14(2), 130–140. https://doi.org/10.3109/17549507.2011.632025

Widaman, K. F. (2006). Missing data: What to do with or without them. In K. McCartney, M. R. Burchinal, & K. L. Bub (Eds.), *Best practices in quantitative methods for developmentalists*. (pp. 42–64). Boston, MA: Blackwell.

Bilingual Turkish-Swedish children's understanding of MAIN picture sequences

Individual variation, age, language and task effects

Ute Bohnacker[1], Buket Öztekin[1] and Josefin Lindgren[1,2]
[1]Uppsala University / [2]Leibniz-Centre General Linguistics (ZAS)

This study investigates story comprehension in 100 bilingual Turkish-Swedish children aged 4 to 7 years, growing up in Sweden with Turkish as their home language and Swedish as the societal language. Information about language development, exposure and other background factors was obtained via parental questionnaires. In both languages, children told two picture-based stories from the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2012, 2019) and answered standardised comprehension questions that probe inferencing of goals and emotions of story characters.

Overall comprehension scores and response accuracies to individual questions were calculated. Story comprehension was compared across ages, languages and tasks, and related to performance on Turkish and Swedish vocabulary tasks (Cross-Linguistic Lexical Tasks, CLT, Haman et al., 2015). A qualitative analysis explored characteristics of the MAIN picture sequences and the type of inference required to score correct on comprehension questions. Overall comprehension scores did not differ between Turkish and Swedish at group level. Comprehension scores increased significantly with age in both languages. This increase was steeper in the majority language Swedish. Younger children (age 4–5) often performed well in Turkish, whilst more older children (age 6–7) performed well in Swedish. In both languages, older children reached relatively high scores, but did not yet master all aspects of inferential story understanding as probed by MAIN. Regression models indicate that a large part of the variance in story comprehension can be explained by age and expressive vocabulary knowledge (CLT) in the respective language. Individual case studies of exceptionally poor story comprehenders vs. high performers also suggest that story comprehension and vocabulary skills are linked, but moreover that MAIN comprehension is influenced by language input and use in and outside the home.

An interesting task effect was found, indicating that the comprehension measure for the MAIN Cat and Dog picture sequences is easier than for Baby Birds/Baby Goats – even when they are administered in the very same mode.

https://doi.org/10.1075/sibil.61.04boh

The task influenced children's comprehension performance more than the language of testing did. Turkish and Swedish showed the same overall response patterns, with very high vs. low performance on certain individual questions. We argue that due to subtle differences in the pictorial stimuli, parallel and seemingly identical comprehension questions require inferences with rather different levels of difficulty. Comprehension scores should therefore not be straightforwardly compared across MAIN tasks.

Keywords: inferencing, story comprehension, Swedish, Turkish, vocabulary

1. Introduction and background

This chapter investigates the development of story comprehension in Turkish-Swedish bilingual children aged 4 to 7, in their heritage (or home) language Turkish and in the societal language Swedish.

There is a growing literature internationally on the narrative abilities of bilingual children, but this research largely focuses on story production, not story comprehension. Here, the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012, 2015) offers new possibilities. MAIN contains 4 carefully designed, structurally parallel fictional picture sequences (the Cat, Dog, Baby Birds and Baby Goats stories) and a standardised protocol to assess both narrative production and comprehension in children aged 4–10. Bilingual children can thus be tested in both their languages with comparable stimulus materials, without reusing the same story. This minimises practice/training effects and/or boredom, which easily occur when the child is tested repeatedly on the same material (Pavlenko, 2009; Klop, Visser, & Oosthuizen, 2012). The present study investigates story comprehension with MAIN.

But what is story comprehension? Comprehension of any type of discourse means that the comprehender constructs a coherent representation of what the discourse is about in her/his mind (e.g. Kintsch & van Dijk, 1978; Stein & Glenn, 1979; Trabasso, Secco, & van den Broek, 1984). Stories are a particular type of discourse, and one of their fundamental characteristics is that they have a causal event structure. Thus, in story comprehension, the comprehender constructs a coherent (and meaningful) mental representation of what the story is about; and in order to do so needs to identify and infer meaningful relations, including the motivational and causal relations between events. According to a number of researchers, this representation takes the form of a mental network that mirrors the causal relations between events that the comprehender has recognised or inferred (e.g. Trabasso et al., 1984; Trabasso & Nickels, 1992).

Comprehenders of stories construct their mental representations on the basis of two things: firstly, the information provided in the story (which may be presented via different media, a written text, an oral narrative, an audio-visual narrative (e.g. film), or purely visually in a silent film, a nonverbal video clip, a picture sequence or a wordless picture book (Ellis Weismer, 1985; Bishop & Adams, 1992; Bishop, 1997, Chapter 7; Paris & Paris, 2003; Lynch, van den Broek, Kremer, Kendeou, White, & Lorch, 2008; Hayward, Schneider, & Gillam, 2009)), and secondly, the comprehender's background knowledge and expectations arising from this knowledge.

Whilst even very young children are able to identify and infer meaningful relations between events to some degree and create network representations of these events (Thompson & Myers, 1985; Lynch et al., 2008), their story comprehension is generally not as developed as that of older children and adults. This is hardly surprising, as children with every year of their lives accumulate new experiences and knowledge of facts about the world, as well as social knowledge of how humans interact. With increasing experience and background knowledge, they can then make more efficient use of their attentional and working memory capacities in story comprehension (cf. van den Broek, Kendeou, Kremer, Lynch, Butler, & Lorch, 2005).

Story comprehension is a prerequisite for story production: If children do not understand cause-effect relationships, plotlines, intentions, thoughts and feelings of story characters, they will not be able to convey this information to the listener during storytelling either (Stein & Glenn, 1979; Shapiro & Hudson, 1991, p. 115; Trabasso & Nickels, 1992; Astington & Pelletier, 2005, p. 327). In the world of stories, autonomous agents (i.e. story characters) have intentions and goals and act and react in relation to these goals. To understand the events depicted or described, comprehenders often assume the story character's perspective, mentally simulating their internal states (Mar, 2004. p. 1416).

In story comprehension and storytelling, children must thus make inferences about the internal (or mental) states of others (e.g. Bishop, 1997; Letts & Leinonen, 2001; van den Broek et al., 2005). This requires theory of mind, i.e. the awareness of mental states and "the ability to use this awareness in interpreting, explaining and predicting the behavior of one self and others" (Astington & Pelletier, 2005, p. 313). Here, a number of studies have found that children below the age of 8–9 rarely make explicit mention of the internal states of characters when telling fictional stories, whilst older children and adults do express them (e.g. Shapiro & Hudson, 1991; Trabasso, Stein, Rodkin, Munger, & Baughn, 1992; Berman & Slobin, 1994; Ukrainetz, Justice, Kaderavek, Eisenberg, Gillam, & Harm, 2005). But even if young children rarely mention internal states in their storytelling, the very same children are able to answer inferential questions that directly probe these internal states, e.g. the goals and emotions of story characters. This has mostly been shown for monolingual English middle-class children aged 4–6 (Stein & Glenn, 1979; Trabasso et al., 1992; Lynch & van den Broek, 2007; Hayward et al., 2009; Tompkins, Guo, & Justice, 2013).

MAIN (Gagarina et al. 2012, 2019) includes both storytelling and story comprehension tasks with picture sequences as a base. Story comprehension is assessed via probe questions that require the child to make inferences from the pictures and to verbalise this understanding.

As MAIN is still a relatively new tool, only around 10 publications have reported results for story comprehension in MAIN (Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016 for Dutch; Bohnacker, 2016, for English-Swedish; Kapalková, Polišenská, Marková, & Fenton, 2016, for Slovak-English; Maviş, Tunçer, & Gagarina, 2016, for Turkish-German; Roch, Florit, & Levorato, 2016, for Italian-English; Rodina, 2017 for Russian-Norwegian; Lindgren, 2018 for Swedish, German-Swedish and Turkish-Swedish; Lindgren, 2019, for Swedish; Bohnacker & Lindgren, in press, for Swedish and English-Swedish; Otwinowska, Mieszkowska, Miałecka-Pikul, Opacki, & Haman, 2018, for Polish and Polish-English); this number will be greatly boosted by the contributions in the present book volume. The above studies have found story comprehension on MAIN to be developmentally ahead of story production, which confirms earlier work that used different materials (Hayward et al., 2009; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012). However, to date, only very few studies (Bohnacker, 2016; Lindgren, 2018, 2019; Bohnacker & Lindgren, in press) have investigated story comprehension in MAIN in any depth, compared comprehension across the different stories or explored answer patterns to individual comprehension questions.

So far, a handful of studies have explored how MAIN story comprehension develops with age in bilinguals (Bohnacker, 2016; Maviş et al., 2016; Roch et al., 2016; Lindgren, 2018). Based on their cross-sectional data, significant increases in comprehension scores from age 4 (or 5) to age 6 and 7 emerge. It is of course only to be expected that story comprehension improves with age – alongside increasing cognitive maturity and inferencing abilities, advances in processing and children's expanding world knowledge. As yet, no norm-referenced data exist for comprehension in MAIN. In a study of 124 children aged 4–6 (72 Swedish monolinguals and 52 English-Swedish bilinguals from high-SES backgrounds), Bohnacker and Lindgren (in press) have made a first attempt to establish benchmarks. They put forward suggestions for what we can expect of a typically developing child at age 4 and age 6, concerning overall comprehension scores for the different MAIN stories, and for performance on individual comprehension questions. It remains to be seen whether the benchmarks suggested by Bohnacker and Lindgren (in press) can be shown to hold for children of other language combinations and backgrounds as well.

Another finding concerning story comprehension on MAIN has been that the bilingual children studied so far score similarly in their two languages, at least at age 6–7 (Bohnacker, 2016; Kapalková et al., 2016; Roch et al., 2016; Lindgren, 2018; Bohnacker & Lindgren, in press). For younger bilingual children, it has been suggested that there may be more variation in the comprehension scores in the two languages due to effects of limited, uneven language exposure, though this issue has not been systematically studied yet. Inasmuch as the above studies report details on the background of their participants, the majority come from mid- to high-SES homes, with high parental education levels and regular shared book reading and storytelling activities. Such background characteristics may well affect the children's performance and pace of development in story comprehension, but this is uncharted territory. Similarly, no studies have yet been carried out that investigate the potential relationship between children's general linguistic skills (such as vocabulary knowledge) and their story comprehension in MAIN. The present study and other contributions in this volume address some of these knowledge gaps.

Assessment of children's story comprehension should not be conflated with assessment of memory skills. Story comprehension is not the ability to recall as many story events or facts as possible, but the identification of causal, hierarchical and thematic relations between events and groups of events, resulting in an understanding of the overall plot or point of the story (van den Broek et al., 2005, pp. 118, 126). However, in a number of studies and materials (e.g. the Bus Story Test, Renfrew, 1969), the child first listens to a story, with or without pictorial support, and then has to retell or recall the story. Recall is taken as an indirect measure of story comprehension; however, as the child has to hold large chunks of aurally presented information in working memory, this type of 'story comprehension' assessment in effect becomes a memory task (Boudreau, 2007; Dodwell & Bavin, 2008; Bohnacker & Lindgren, in press). A similar conflation occurs when the child first listens to the story and is then asked comprehension questions. To score correct on such questions, the child must recall details that were previously mentioned or shown, or, if the question is an inferential one, make inferences. Yet if the to-be-recalled facts are only presented aurally, story comprehension is again heavily intertwined with memory. In contrast, if the stimulus pictures are kept in full view of the child and can be referred to during comprehension questions, working memory is taxed far less. Sometimes, story presentation is in the visual modality only (e.g. Ellis Weismer, 1985; Bishop & Adams, 1992; Paris & Paris, 2003; Hayward et al., 2009; Gagarina et al., 2012, 2019). The child does not hear the story first but studies a picture sequence, and is then asked comprehension questions that assess inferencing abilities, whilst being able to look at and refer to the pictures. In the present study, this latter method is used.

Yet even when comprehension questions are asked with the above method, they still do not directly assess the child's understanding of a story, but rather the ability to *verbally express this understanding*. Answering comprehension questions requires expressive language skills, i.e. a certain level of proficiency in the language that story comprehension is tested in. A child might understand the story but, due to limited language proficiency, still not be able to convey this understanding to the listener/experimenter. As MAIN assesses the understanding of goals and internal states (which are *not* depicted in the stimulus pictures but must be inferred), the MAIN author group, when developing the MAIN materials and protocol, did not think it feasible to probe such aspects of story comprehension in a nonverbal manner. It is not possible to explain how story characters feel and think and why they feel and think that way by purely nonverbal means such as pointing to a picture or drawing a happy or sad face.

As far as we know, there are virtually no published studies that systematically investigate the role which general language proficiency, such as vocabulary knowledge, plays in bilingual children's story comprehension.[1] The present study therefore includes an analysis of how lexical knowledge, as measured on a vocabulary task, predicts (or influences) children's story comprehension scores on MAIN. This is done for both languages of our bilingual participants. We are not aware of any publications that systematically investigate how background characteristics, such as language exposure, preschool attendance, home-language use, parental education and cultural family habits concerning book reading and storytelling in the home, are related to bilingual children's narrative comprehension.[2] The present study therefore explores such potential links by including case studies of individual children.

1. Westerveld (2014), in a study of emergent literacy in 18 bilingual Samoan-English 4-year-olds in New Zealand, briefly considers the relation between vocabulary and narrative listening comprehension.

2. For monolingual children, a few studies have explored the relation between narrative comprehension and home literacy activities and/or vocabulary knowledge. For instance, Sénéchal and LeFevre (2002) investigated monolingual Canadian English (upper) middle-class children aged 4 to 8 and found correlations between measures of storybook exposure in the home, receptive vocabulary (PPVT, Dunn & Dunn, 1997) and narrative listening comprehension. The authors did not provide details on how narrative comprehension was assessed. Lynch et al. (2008) studied monolingual U.S. English middle-class children at age 4 and 6 and found correlations between their receptive vocabulary scores (PPVT) and narrative listening comprehension (probed recall of events). Westerveld, Gillon, and Boyd (2012, p. 136) assessed monolingual English preschoolers age 4 to 5 (mid-to-high SES) and found that receptive vocabulary (PPVT) correlated with response accuracies to factual and inferential comprehension questions asked after listening to a story. Lepola et al. (2012, p. 269), in a study that mainly dealt with listening comprehension

The paper proceeds as follows. Section 2 states our aim and research questions. Section 3 describes the methodology, including a detailed description of common background characteristics of our participants. Results are reported in Section 4. As we explore story comprehension in bilingual children from several different angles, Section 4 contains the following subsections: 4.1 reports on overall story comprehension scores and differences between languages (Turkish, Swedish), tasks, and age groups (4-, 5-, 6-, and 7-year-olds). Section 4.2 explores individual differences, including exceptionally low-scoring and high-scoring children and their backgrounds. In Section 4.3, the relation between story comprehension scores and vocabulary knowledge is investigated. Section 4.4 explores task differences further, by looking at response accuracies and answer patterns for individual comprehension questions in MAIN and possible explanations for these. Finally, Section 5 contains a discussion and conclusion.

2. Aim and research questions

The aim of this study is to investigate the story comprehension skills of Turkish-Swedish bilingual children aged 4 to 7 in both languages with MAIN (Gagarina et al., 2012, 2019). Overall comprehension scores and response patterns to specific comprehension questions are analysed for 4 different narrative comprehension tasks (2 in each language). The following research questions are asked:

1. How does story comprehension, as measured with MAIN, develop with age in Turkish and Swedish?
2. Is there a difference in story comprehension between the languages (Turkish, Swedish)?
3. Do children perform differently on the different MAIN stories (Cat, Dog, Baby Birds and Baby Goats)?
4. Are there differences in response accuracy between individual comprehension questions, and if so, why?
5. Is story comprehension influenced by lexical knowledge, as measured on a vocabulary task?
6. Exploring individual variation, what are the characteristics of high-scoring and low-scoring children (e.g. general language proficiency, language exposure and socio-economic background)?

in 4-to-6-year-old Finnish-speaking children of highly-educated mothers, found a side effect, namely that child performance on a vocabulary definition task correlated with response accuracy to inferential questions about a picture story from Paris and Paris (2003).

3. Method

3.1 Participants

The participants were 100 4-to-7-year-old Turkish-Swedish bilingual children growing up in Sweden. The children had Turkish as their home language, whilst Swedish was the language of schooling and society at large. Basic inclusion criteria were: (i) ability to speak both Turkish and Swedish, (ii) 4–7 years, and (iii) no diagnosis of language impairment. The majority of children were reached by contacting around 200 preschools and schools in urban areas of eastern Central Sweden. Other participants were recruited through personal contacts, municipal heritage-language teachers of Turkish, places of worship and community centres. In the end, participating children came from 50 different preschools and schools in the Greater Stockholm area and two larger nearby cities.[3]

The participants were part of a larger multilingualism research project (Bohnacker, 2013–2019). In the present chapter, only those children are included for whom we have both narrative comprehension and vocabulary data. These 100 children (aged 4;0–8;1) were divided into four age groups: 4-year-olds, 5-year-olds, 6-year-olds and 7-year-olds. Table 1 provides a breakdown by age and sex.

Table 1. Participants

	4 years	5 years	6 years	7 years	Total
N	24	22	26	28	100
Girls/boys	12/12	13/9	15/11	14/14	54/46
Mean age	4;5	5;5	6;5	7;5	6;0
Age range	4;0–4;11	5;0–5;11	6;1–6;11	7;0–8;1*	4;0–8;1

Note. * Three children in the 7-year-old group had just turned 8 years at the time of testing.

Informed parental consent was obtained in writing. Families and children could terminate their participation in the study at any time. Detailed information about the child's language development, patterns of language use in the family, the parents' language and educational backgrounds etc. was gathered via a 5-page paper-and-pencil questionnaire. Parents filled in the questionnaire in the language of their choice; and from some parents who disliked writing, information was gathered via telephone interview in Turkish. Parental questionnaire data are available for all 100 children. The data were anonymised.

3. Random sampling from the national population register was not possible, as no statistics are kept on whether or not a resident of Sweden speaks Turkish (or any other language).

According to the parents, all children had typical language development and none had been diagnosed with language impairment or a neurological disorder. Children who were exposed to an additional third language in the home, such as Kurdish, were not excluded, since they did not generally perform differently from the strictly *bi*lingual children on our Turkish and Swedish language measures, and since they form an integral part of the Turkish-Swedish population.

All children were living in Sweden at the time of data collection and nearly all, 92% (92/100), had done so from birth. By contrast, only 22% of the parents had been born and grown up in Sweden. Most parents (72%) had been born and grown up in Turkey and migrated to Sweden as adults, and many of the remaining parents had been born in Turkey but moved to Sweden as children. For most children then, both parents were first-generation immigrants from Turkey or one first-generation and one second-generation parent. First-generation immigrant parents had lived in Sweden from anywhere between 0.6 to 41 years. A large majority of parents (74%) stated that Turkish was their native language (L1). 4% reported Swedish as their L1, and 2% considered both Turkish and Swedish as their L1s. 20% of the parents reported a L1 different from Turkish or Swedish; in nearly all cases, this language was Kurdish. This means that very few children (8/100) had a parent who was a native speaker of Swedish.

With Turkish so strongly represented as the parental L1, virtually all children in the study (96%, 96/100) had received Turkish input from their parents since birth. Only for 4 children did exposure to Turkish start after 1;0 (and before 3;0). The reason for this was that these children were first exposed to Kurdish, as their parents had Kurdish as an L1.

Onset of exposure to Swedish varied more than for Turkish. Most children (83%, 83/100) had started to hear Swedish regularly before age 3;0 (with an even distribution for onset of Swedish between birth and 1;0 (27%), 1;0 to 2;0 (26%), and 2;0 to 3;0 (28%)). For 17% of our participants, exposure to Swedish began after age 3; mostly between 3;0 and 4;0. Only very few participants were first exposed to Swedish after age 4; these were exclusively children born in Turkey who had later moved to Sweden.[4]

We explored language use in the home through a number of questionnaire questions, including the parents' language(s) spoken with each other and to the child, the child's language spoken to the parents and to the siblings, and language use between siblings. Details can be found in Öztekin (2019). A majority of parents reported that they spoke almost only or mostly Turkish to their child (65%)

4. We did not exclude children born in Turkey (with at least two years exposure to Swedish), since they did not generally perform differently on our measures of Swedish and Turkish than the children born in Sweden.

and to each other (61%). There were no couples in the sample who spoke only/mostly Swedish to their child (0%) and only two couples who spoke only/mostly Swedish to each other (2%). The remainder spoke a mix of Turkish and Swedish or another language. The majority of children (64%) were reported to speak almost only/mostly Turkish to both parents; but notably, 12% children spoke only/mostly Swedish to their parents, suggesting that a generational shift of language preference is beginning to take place. Between siblings, only/mostly Swedish was reported for 15% of the children, 46% communicated with their siblings in both Swedish and Turkish, while only 30% did so only/mostly in Turkish, which is very different from parental language use.

The questionnaire also asked parents to estimate the proportion of daily language input to the child in Turkish and Swedish. Parents estimated this on a scale ranging from 5 percent to 95 percent for each language. The majority of children (65%) were reported to receive approximately equal amounts of Turkish and Swedish during the day (40:60, 50:50, 60:40). For 20 children (20%), parents reported daily input to be at least 80% Swedish; for 9 children (9%), at least 80% Turkish input was reported. The few remaining children had other input distributions, including a third language.

All children were active bilinguals. 6% of the children (6/100) also spoke a third language: Kurdish (Kurmanji), Zaza (Dimili), German or English, but for none of them was this third language rated strongest or as strong as Turkish or Swedish. A few other children did not speak a third language but were exposed to it; this mainly occurred in families where the parents had Kurdish as their L1 and spoke a mix of Turkish and Kurdish to each other.

The majority of parents rated their children's receptive and expressive language skills in both Turkish and Swedish as 'very good' or 'good', though some rated one language as 'so-so' or even 'bad'. For 39% of the children, the parents stated that the child's language skills were higher in Turkish than in Swedish. A nearly identical number (38%) was said to have equal language skills in both languages. For a smaller but substantial group (24%), parents rated their children to be more proficient in Swedish. It is our impression that parents with low proficiency in Swedish may have overestimated their child's level of Swedish (cf. Koivistoinen, 2012, p. 33; Tuller, 2015). Most parents (81%) considered Turkish and Swedish to be equally important for their child (this also included 7 cases where a third language was rated equally important alongside Turkish and Swedish). 15% of the parents considered Turkish most important, 3% Swedish, and 1% Kurdish most important.

Virtually every child attended institutional childcare for a major part of the day, as is typical in the Swedish context. Most children had started attending preschool between age 1 and 2, as is usually the case in Sweden; the exception being mainly Turkey-born children who had moved to Sweden. Early (and extensive) attendance of Swedish daycare presumably has a positive impact on the development of

Swedish language skills.[5] Most children (83%) were reported to regularly attend (pre)school for 30–48 hours per week, 10% attended 24–26 hours, and 7% only 6–20 hours/week. All schools and preschools were run in Swedish, though on closer inspection (during our data collection visits) a number of them did afford the child opportunities to communicate in Turkish, as other children and/or a staff member were also Turkish-speaking.

The questionnaire also queried language-related activities for both languages in and outside the home, such as storytelling, joint book reading, singing, media consumption and heritage language instruction. 55% of the children attended Turkish heritage language classes (ca. 40–60 min/week). 75% regularly consumed Turkish media, which was more than Swedish media. 50% were exposed to Turkish books in the home, which was more than to Swedish books. For 65% of the children, parents stated that they read books in Turkish for the child at least once a week. Regular parental storytelling in Turkish was reported for 59% of the children, singing or listening to songs in Turkish for 64%. These activities were on average more frequent in Turkish than in Swedish. However, the frequencies of such activities varied greatly throughout the sample. Interestingly but perhaps not surprisingly, parents with very low educational levels reported very little literacy-related language activities with the child in the home. Parents with L1 Kurdish (or Kurdish and Turkish) rarely reported any of the above language-related activities for Turkish.

Many participants lived in linguistically and culturally diverse, socio-economically disadvantaged urban areas. The individual children came from a wide variety of socio-economic backgrounds, both concerning parental occupations and education, where all levels from non-completed primary school to doctorate were represented. This variation can be considered typical of Turkish-speaking families in Sweden. We used parental education as a proxy for family SES (socio-economic status). Questions about parents' education were asked and answered separately for each parent. Answers were coded following the United Nations ISCED 2011 classification (UNESCO Institute for Statistics, 2012) and assigned to one of 9 levels ranging from 0 (= only early childhood education) to 8 (= doctoral studies). The mean educational level was 3.4 (out of 8), and 3 (= completed secondary education) was particularly common. The educational levels of both parents were averaged and categorised as 'low SES' in the Swedish context for ISCED levels 0 to 3 (i.e. early childhood education up to secondary school) vs 'high SES' for ISCED levels 4 to 8 (i.e. completed upper secondary education plus at least some tertiary education, or more). In a few cases, information about education was missing; since this mainly occurred with parents with very low status occupations, it is likely that their

5. Preschool in Sweden is normally attended until age 5 or 6, after which children attend one year of 'preschool class' (*förskoleklass*) to prepare them for school proper. At age 7, children start grade 1 of primary school.

educational level was low. For those children that we did have parental education information for (*N* = 95), more fell into the low SES category (64%, 61/95) than into the high SES category (36%, 34/95). The majority of parents did thus not have any tertiary education. The age groups (4-, 5-, 6-, and 7-year-olds) did not differ from each other with respect to SES, nor for any of the other above-mentioned background factors, except Turkish heritage-language class attendance (which was more frequent in the 6- and 7-year-olds than the younger children).[6]

Despite considerable diversity, the background of the majority of our participants appears to match what has been observed for families with Turkish as a heritage language elsewhere, namely that endogamy and a continued influx from Turkey revitalise and keep up the use and transmission of the home language to the next generation. Turkish is regarded as important alongside the majority language Swedish, which the children are exposed to from an early age via preschool. However, in contrast to several other Western European countries, the migration history from Turkey to Sweden is not dominated by labourers, but by political refugees and/or family members (via marriage and family reunification), with a variety of educational backgrounds. This includes many Kurdish/Turkish speakers, some of whom may choose bring up their children in a trilingual environment.

3.2 Materials

3.2.1 *MAIN*

This study focuses on children's story comprehension skills, as measured with MAIN (Gagarina et al., 2012, 2015, 2019). MAIN consists of 4 picture sequences with 6 pictures each. Each of these stories is made up of three episodes. Episodes are a chronologically ordered group of events within a larger narrative that are conceptually connected to a specific goal of a story character (Stein & Glenn, 1979). Each episode consists of an internal state as initiating event, goal, attempt, outcome and internal response of the characters. Two of the four picture sequences, the *Cat* and *Dog* stories, are parallel in plotline, story grammar and length. Both Cat and Dog have three main characters; the only difference is the characters and objects in the stories. (Our participants received one of these two stories in Turkish and the other one in Swedish.) The other two picture sequences, the *Baby Birds* and *Baby Goats* stories, are also parallel in plotline, story grammar and length. Both consist of three episodes and have five animal characters each, making them more complex

6. The older children's higher attendance of Turkish classes is probably due to the fact that municipalities in Sweden are obliged to offer heritage language instruction for primary-school pupils, whilst there is nowadays no such legal requirement for preschoolers any more.

than Cat and Dog. Baby Birds and Baby Goats are both about a family (of birds/goats), an attacker, and a rescuer. (Again, our participants were administered one of these two stories per language.) In what follows, *Cat/Dog* will be referred to as one narrative task, and *Baby Birds/Baby Goats* as the other narrative task.

There is a standardised procedure for administering MAIN (see Gagarina et al., 2012, 2019, and the introductory chapter of this book volume for details). The child tells the story and is then asked ten comprehension questions about the internal states and goals of the protagonists. During storytelling, only the child can see the pictures, but when the comprehension questions are asked, the pictures are in full view of both child and experimenter, for joint visual attention. The questions require the child to make inferences from the pictures and state why s/he has come to a certain conclusion. The questions thus assess inferencing, i.e. how well the child is able to interpret physical and psychological (emotional, motivational) cause-effect relationships and recognise characters' goals, the reasons for these goals and reactions following attempts to reach the goals. Importantly, the questions used for the four stories are all structured the same way, asking for the same types of essential story information. This enables comparisons across stories (in contrast to many earlier studies of comprehension that asked an ad hoc set of questions that were tied to particular stories but did not attempt to collect the same kind of information across stories).

Table 2 exemplifies the comprehension questions for the Dog story, and Figure 1 shows the Dog picture sequence in small scale.[7]

Table 2. The ten comprehension questions in MAIN Dog

Questions	Example
D1. Episode 1 Goal	Why does the dog jump/leap forward? (point to picture 1–2)
D2. Episode 1 IS	How does the dog feel? (picture 3)
D3. Episode 1 IS rationale	Why do you think that the dog is feeling …? [insert answer D2]
D4. Episode 2 Goal	Why does the boy jump up? (picture 5)
D5. Episode 2 IS	How does the boy feel? (picture 6)
D6. Episode 2 IS rationale	Why do you think that the boy is feeling …? [insert answer D5]
D7. Episode 3 Goal	Why does the dog grab the sausages? (picture 5)
D8. Episode 3 IS Theory of Mind	Imagine that the boy sees the dog. How does the boy feel? (picture 6)
D9. Episode 3 IS rationale (ToM)	Why do you think that the boy feels …? [insert answer D8]
D10. Overall plotline (ToM)	Will the boy be friends with the dog? Why?

Note. IS = internal state. ToM = Theory of Mind.

7. All four picture sequences are shown in Figures 8–9.

Figure 1. Small-scale copy of the MAIN Dog picture sequence (Gagarina et al. 2012) (reproduced with permission from the publisher)

For all MAIN stories, three questions (D1, D4 and D7) probe understanding of the goals of the main characters in the three episodes. Questions D2 and D5 query the internal states of the characters in Episode 1 and 2. D8 queries a character's internal state in Episode 3, but does so with a theory of mind question, testing the child's understanding by a 'what if' scenario that does not actually happen in the story. Questions D3, D6 and D9 are follow-up questions to the preceding questions about a character's feelings. These follow-up questions ask *why* a character feels a certain way. If the child does not answer the preceding question correctly, the follow-up question is not asked according to MAIN protocol. The final question, D10, tests whether the child can infer meaning about the story as a whole (overall plotline). It is made up of both a yes/no question and a 'why' question, and only scored as correct if the child answers both questions correctly.

3.2.2 *Vocabulary: CLT*

While the focus of the present study is on narrative comprehension, we also wanted to see whether children's lexical knowledge had an impact on MAIN comprehension. To get an estimate of vocabulary comprehension and production, comparable picture-based vocabulary tasks were used in both languages, namely the Turkish and Swedish versions of the Cross-linguistic Lexical Tasks (CLT, Haman, Łuniewska, & Pomiechowska, 2015; see <http://psychologia.pl/clts/>). Each CLT version consists of 4 parts (noun comprehension, verb comprehension, noun production and verb production), and contains 120 vocabulary items, 30 in each part (plus 2 practice items). Comprehension is tested via picture selection, where the child has to point to the one correct picture out of four in response to prompts such as "Where is the spoon?" (for nouns) or "Who is painting?" (for verbs). Vocabulary production is tested via picture naming, where the experimenter asks the child to name a depicted object ("What is this?" for noun production) or action ("What is she doing?" for verb production).

3.3 Procedure

Each child told two MAIN stories per language and answered the comprehension questions for these stories. Each child also did a vocabulary task (CLT) in each language. Turkish and Swedish were tested on different occasions with at least four days in between. On average, the time between sessions was 13.6 days (*SD* = 7.9, range 4–36 days).[8] The order of the languages was counterbalanced: Half of the children in each age group were tested in Turkish first and the other half in Swedish first. Children were tested in Turkish by the second author or by one of two research assistants, all native speakers of Turkish. The children were tested in Swedish by the first or the third author or one of five other members of the Swedish-speaking project team, all native or near-native speakers of Swedish. Before testing, the experimenter met either the teacher of the child at (pre)school or the child's parent(s) in the home. As a rule, testing took place in a quiet room in the home or (pre)school, with the child alone. The experimenter only spoke to the child in the language of testing and acted as if s/he did not understand the other language. The two narrative tasks in each language were carried out within one session as part of a larger test battery. Cat/Dog was always done first, followed by the CLT, which in turn was followed by Baby Birds/Baby Goats. The child never received the same story twice. The stories were counterbalanced within each age group and language. The procedure for all narrative tasks was as follows: First the child told the story and then answered comprehension questions.[9] All children were asked the 10 standardised questions from the Turkish and Swedish MAIN. The CLTs were administered according to standard procedure (Haman et al., 2015). The order of the four parts of the CLT was counterbalanced within each age group. CLT child responses were documented on paper. The entire session was audio and video recorded. Children were rewarded with stickers and a diploma.

3.4 Scoring

3.4.1 *Scoring of the MAIN narrative comprehension questions*

Responses to the MAIN comprehension questions were transcribed verbatim.[10] Responses were scored as 1 (correct) or 0 (incorrect), so with 10 questions per

8. Gagarina et al. (2012) recommend a time lapse of 5–7 days between the two language sessions for MAIN, but this was not always practicable due to illness, holidays or unexpected events.

9. Note that all MAIN tasks were done as storytelling, even though Cat/Dog had originally been created to be used for retelling or model story (Gagarina et al., 2012). In the present study, testing mode (storytelling followed by comprehension questions) was kept constant, so that all tasks were equivalent and comprehension results could be straightforwardly compared without testing mode being a confound.

10. The MAIN stories told by the children were also transcribed and analysed, but are not dealt with in the present study. For story production results, see Lindgren (2018) and Öztekin (2019).

story, the maximum score is 10 points. A total of 4,000 comprehension questions were to be asked (100 children x 2 languages x 2 stories x 10 questions). All children (except three) received four different MAIN comprehension scores (two in Turkish, two in Swedish), one for each story. These are their overall comprehension scores.

Comprehension scores were missing in three cases: In Turkish, one 5-year-old refused to answer the questions on Baby Goats. In Swedish, the experimenter skipped more than three questions in one of the stories of two 4-year-olds, so no reliable comprehension score could be calculated. Thus, 3,970 questions remained. Due to experimenter error, 19 out of 1,990 Turkish questions were not asked, resulting in 0.9% missing data for Turkish. In Swedish, the experimenter skipped or forgot to ask 23 out of 1,980 questions, which makes 1.2% missing data in Swedish. As the child did not get the chance to hear and answer these questions, it would have been unfair not to award any points. Following common practice when there is no more than 2% missing data (Widaman, 2006), sample mean substitution was used, so that for each missing question, a score identical to the mean for that question for the child's age group was substituted.

All the Turkish answers were scored by the second author. The Swedish answers were scored by the third author and a trained native Swedish research assistant. Unclear cases were discussed with the first author. As child responses often did not correspond to the examples given in the original MAIN scoring protocol (Gagarina et al., 2012), the first author and her research group developed detailed scoring guidelines (Bohnacker, 2018b) that included general scoring principles and a large number of real-life answers (with the score and a rationale for the score). Scoring principles and decisions were based on multiple rounds of discussion of the answers of 286 mono- and bilingual children. In the current study, these guidelines were closely followed, and a trained research assistant checked all answers for consistency in scoring. Answers that were incomprehensible, semantically vague, or grammatically so rudimentary that they left lots of room for interpretation were not awarded points. In line with the MAIN protocol, purely gestural, nonverbal responses were not awarded points, even though the child might have understood the question.

3.4.2 *Scoring of the vocabulary tasks (CLT)*

Child responses were documented on paper during testing and later checked against the audio and video recordings. Responses were scored as 1 (correct) or 0 (incorrect) by the three authors (native or near-native speakers of Turkish and Swedish) and a trained research assistant. Responses were carefully checked for consistency and accuracy, following CLT scoring guidelines developed by the first author and her research team on the basis of data from 220 monolingual and bilingual children (Bohnacker, 2018a), to complement the short information contained in the CLT test materials.[11] Each child received 4 CLT vocabulary scores (with maximally 60 points each), two in Turkish and two in Swedish, one each for comprehension and production.[12]

4. Results

4.1 MAIN comprehension scores: Differences between languages, tasks and age groups

As described above, in each language, participants were always tested on Cat/Dog first, and second on Baby Birds/Baby Goats. We will therefore refer to Cat/Dog as MAIN1, and Baby Birds/Baby Goats as MAIN2. Table 3 shows the mean scores for each story for all the children combined.

Table 3. Mean MAIN comprehension scores, all children combined. Max score = 10 points

	MAIN1		MAIN2	
	Cat	**Dog**	**Baby Birds**	**Baby Goats**
Turkish				
N	50	50	52	47
Mean	7.3	7.1	5.5	6.8
SD	2.1	2.7	2.8	2.3
Swedish				
N	51	48	47	52
Mean	7.4	8.1	6.2	6.3
SD	2.2	2.2	2.5	2.8

Note. N = number of children.

Independent samples t-tests showed that there was no difference in the Swedish comprehension of Cat and Dog ($p = .142$), nor between Baby Birds and Baby Goats ($p = .949$). In Turkish, comprehension of Cat and Dog did not differ either ($p = .614$). However, children who were asked about Baby Goats in Turkish performed significantly better than those who were asked about Baby Birds ($t(95.761) = -2.531$, $p = .013$). Since the difference between these stories was significant, we carried out an additional analysis with Story and Age group for MAIN2 (see below).

11. For details on the CLT scoring, see Bohnacker, Lindgren & Öztekin (2016) and Öztekin (2019).

12. Detailed CLT results in both languages are reported elsewhere (Bohnacker, 2020, Öztekin, 2019).

There were significant and positive correlations between the scores for the same task in the two languages (Cat/Dog: $r = .321$, $p = .001$; Baby Birds/Baby Goats: $r = .370$, $p < .001$). Pearson correlations also showed that the Turkish comprehension scores correlated strongly and positively with age (in months) both for MAIN1 ($r = .350$, $p < .001$) and MAIN2 ($r = .379$, $p < .001$). Likewise, the Swedish scores correlated strongly with age in months, both for MAIN1 ($r = .490$, $p < .001$) and MAIN2 ($r = .719$, $p < .001$).

Apart from age (in months), we also analysed differences between the four age groups. Figure 2 shows the mean comprehension scores for the two narrative tasks in each language for each age group, indicating an increase in all the mean scores, though somewhat differently for the tasks and languages. Moreover, on visual inspection, scores for MAIN1 seem to be higher than for MAIN2 for all groups. Scores in the two languages were relatively similar. To test for differences between age groups, tasks and languages, we carried out a repeated-measures (factorial) ANOVA with Task (MAIN1 vs MAIN2) and Language as within-child variables and Age group as between-child variable as well as the interaction between all variables. When an interaction was significant, post-hoc analyses (simple effects analyses with Pillai's trace or univariate analyses depending on the type of interaction effect) were carried out.

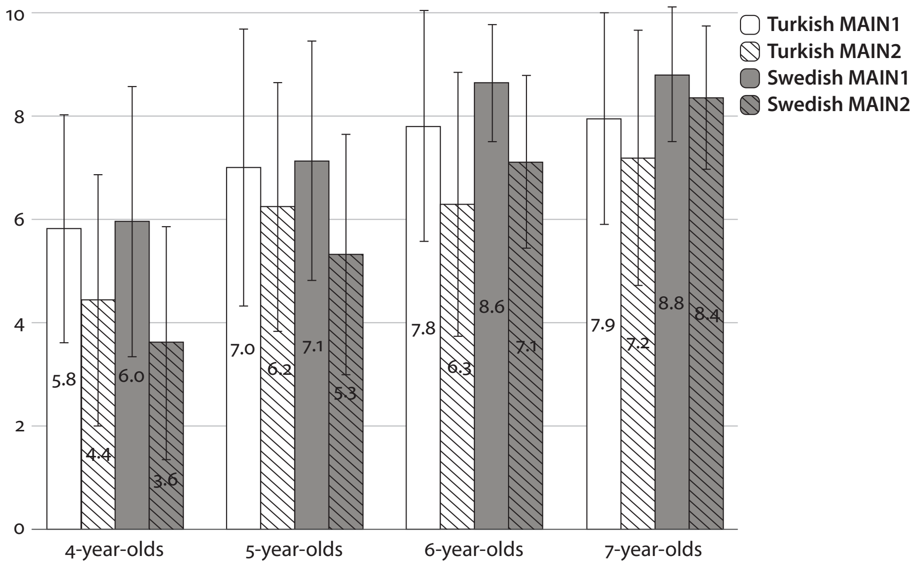

Figure 2. Mean comprehension scores MAIN1 (Cat/Dog) and MAIN2 (Baby Birds/Baby Goats) in Turkish and Swedish, by age group. Max = 10 points. Error bars show ±1 SD.

The analysis showed a highly significant main effect of Task ($F(1, 92) = 75.847$, $p < .001$, $\eta^2_p = .452$), with higher scores on MAIN1 (Cat/Dog) than on MAIN2 (Baby Birds/Baby Goats). The main effect of Language ($F(1, 92) = 1.673$, $p = .199$, $\eta^2_p = .018$) was not significant. The main effect of Age group was highly significant ($F(3, 92) = 20.866$, $p < .001$, $\eta^2_p = .405$). However, the two-way interaction between Age group and Task was also significant ($F(3, 92) = 3.797$, $p = .013$, $\eta^2_p = .110$). This means that the effect of Task may be different for the Age groups *and/or* that the effect of Age group may differ depending on the task. No other interaction effects were significant (*p*s > .10).

Post-hoc tests of the interaction between Task and Age group showed a significant effect of Age group for both tasks (MAIN1: $F(3, 92) = 12.719$, $p < .001$, $\eta^2_p = .293$; MAIN2: $F(3, 92) = 21.880$, $p < .001$, $\eta^2_p = .416$). The larger effect size indicates that the effect was stronger for MAIN2. For MAIN1, pairwise comparisons showed that there were significant differences between the 4-year-olds and both 6- and 7-year-olds (*p*s < .001), and also between the 5-year-olds and the two older groups (*p*s < .01), whereas the difference between the 4- and 5-year-olds just failed to reach significance ($p = .059$), and the 6- and 7-year-olds did not perform significantly different from each other ($p = .722$). For MAIN2, the same type of pairwise comparisons showed significant differences between all age groups (ps ≤ .001), with the exception of the 5- and the 6-year-olds, which was just above the significance level ($p = .057$). The analysis of the effect of Task in the different age groups showed consistently higher scores for MAIN1 in all age groups (*p*s < .05), but with the smallest effect size for the 7-year-olds,[13] indicating that scores on the two tasks become more uniform as children grow older.

For Turkish MAIN2, recall that the children performed better on the comprehension of Baby Goats than Baby Birds. For this reason, we carried out an additional analysis of the Turkish MAIN2 scores, namely an Age group x Story factorial ANOVA. The results showed that there was a significant main effect of Age group ($F(3, 91) = 5.638$, $p = .001$, $\eta^2_p = .157$) and of Story ($F(1, 91) = 7.526$, $p = .007$, $\eta^2_p = .076$). Post hoc tests (Bonferroni-corrected) showed that there was a significant difference in scores only between the youngest and oldest age group, i.e. the 4- and 7-year-olds ($p = .001$).[14] As can be seen in Figure 3, there was no significant interaction between Age group and Story ($F(3, 91) = 0.137$, $p = .937$, $\eta^2_p = .005$).

13. 4-year-olds: $F(1, 92) = 39.218$, $p < .001$, $\eta^2_p = .299$; 5-year-olds: $F(1, 92) = 12.245$, $p = .001$, $\eta^2_p = .117$; 6-year-olds: $F(1, 92) = 28.584$, $p < .001$, $\eta^2_p = .237$; $F(1, 92) = 4.957$, $p = .028$, $\eta^2_p = .051$.

14. The difference between the 4- and 6-year-olds for Turkish MAIN2 scores just failed to reach significance ($p = .064$).

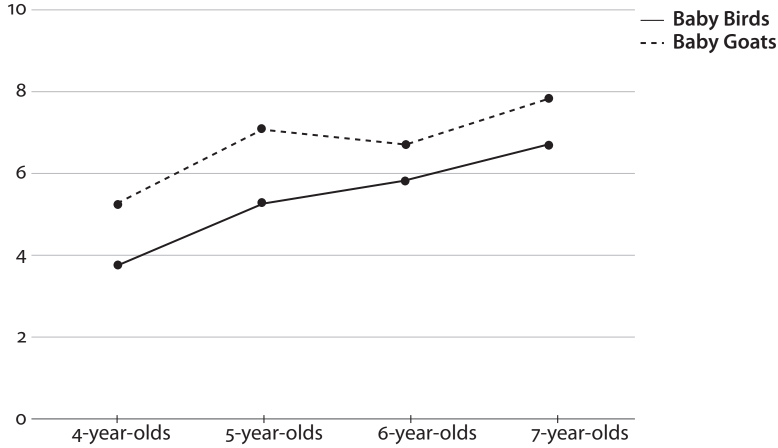

Figure 3. Turkish MAIN2 (Baby Birds and Baby Goats) mean comprehension scores, two-way interaction Age x Story. Max score = 10 points

All age groups had higher mean scores on Baby Goats than on Baby Birds. Possible reasons for the higher performance on Baby Goats will be discussed later.

To summarise, MAIN comprehension scores measurably increased with age in both languages, and comprehension was consistently better for Cat/Dog than for Baby Birds/Baby Goats in both Turkish and Swedish. There was no difference between the two languages in the children's performance.

4.2 MAIN comprehension scores: Individual variation

We found substantial individual variation between participants in MAIN comprehension scores, despite an overall increase with age in months and differences between many of the age groups. Some children scored very low (as low as the minimal score, 0 points) and some very high (at maximum, 10 points). The scatterplots in Figures 4–7 suggest that these very low and very high performers distribute differently for task and language.

In Turkish, very low performers were found amongst both younger and older children, whilst in Swedish, very low performers were restricted to the younger children (4- and 5-year-olds). For Turkish MAIN1 (Cat/Dog, Figure 4), most children scored above 50%, while for Turkish MAIN2 (Baby Birds/Baby Goats, Figure 5), scores were more scattered. In Swedish, the scores clustered more: Very low scores (0–1 points) in Swedish only occurred amongst the younger children, but not in the older age groups. In Swedish, none of the youngest children reached the maximum score, in contrast to Turkish. For Swedish MAIN1, all 6- and 7-year-olds scored well above 50% (Figure 6), and for Swedish MAIN2, virtually all 6- and 7-year-olds did (Figure 7). This suggests that older children scored higher in Swedish than in Turkish in general.

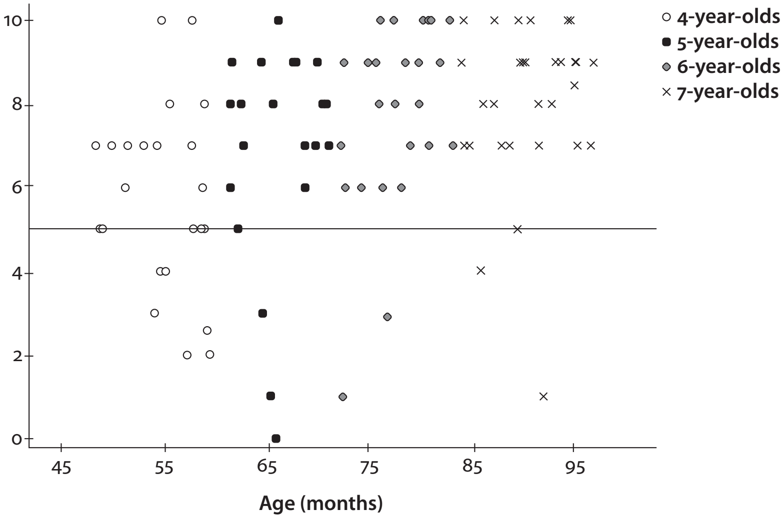

Figure 4. Turkish MAIN1 (Cat/Dog) comprehension scores of each child. Max score = 10 points. The horizontal line indicates 50% (5 points). A dot may represent more than one individual.

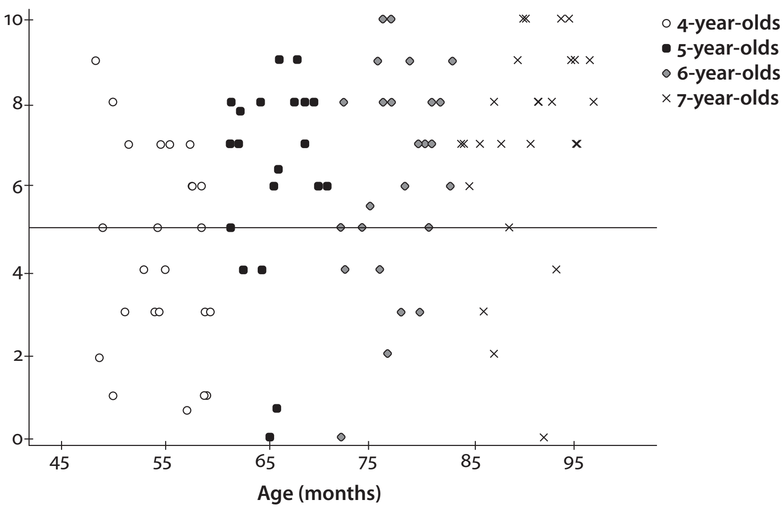

Figure 5. Turkish MAIN2 (Baby Birds/Baby Goats) comprehension scores of each child. Max score = 10. Horizontal line at 50%. A dot may represent more than one individual.

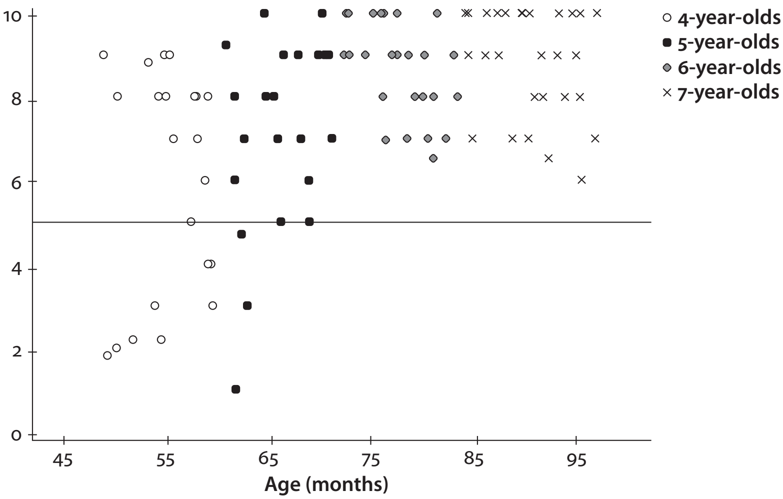

Figure 6. Swedish MAIN1 (Cat/Dog) comprehension scores of each child. Max score = 10. Horizontal line at 50%. A dot may represent more than one individual.

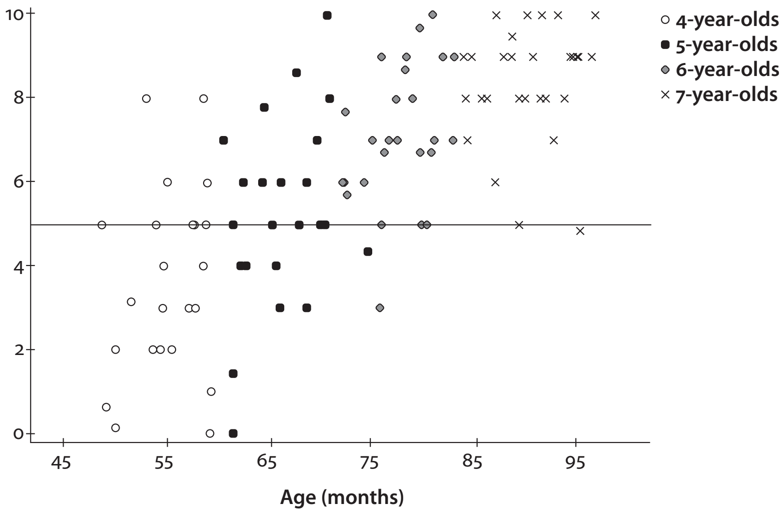

Figure 7. Swedish MAIN2 (Baby Birds/Baby Goats) comprehension scores of each child. Max score = 10. Horizontal line at 50%. A dot may represent more than one individual.

As illustrated by the above scatterplots, certain children stood out with very low or very high scores compared to their age peers. To explore the characteristics of these individual children, we now discuss their MAIN comprehension scores in relation to their vocabulary test results and socioeconomic and language backgrounds. This takes the form of case studies in Sections 4.2.1 and 4.2.2. The relation between MAIN comprehension and vocabulary is explored statistically for all children in Section 4.3.

4.2.1 *Exceptionally low-scoring children*

Two 5-year-olds, two 6-year-olds and one 7-year-old scored extremely low on narrative comprehension in Turkish, with scores of 0–1 points (out of 10). The child BiTur5-21 scored 0 on both MAIN1 and MAIN2 (cf. age group means $M_{MAIN1} = 7.0$, $M_{MAIN2} = 6.2$). This particular child also performed very poorly on the Turkish vocabulary tasks (CLT comprehension score: 26 out of 60; CLT production score: 3 out of 60; cf. age group means $M_{compr} = 53.8$, $M_{prod} = 37.6$). The child's Swedish narrative comprehension scores (MAIN1 = 5 points, MAIN2 = 3 points) also belonged to the lowest in her age group. What else do we know about this child? She was born in Sweden and exposed to Turkish, Kurdish and Swedish in the home. (Note that children exposed to three languages did not generally perform lower than those exposed to two languages.) According to BiTur5-21's parents, the daily input was 80% Swedish. The parents were Turkey-born L1 Kurdish speakers, one of them with very low education and presumably low literacy (ISCED level 1; the other parent: level 3). No Turkish book reading or storytelling activities were reported, but occasional activities in Swedish. The parents considered Swedish to be the child's stronger language. They expressed concern about her development in Turkish, adding that the child had some family members who had difficulties with speech and language. BiTur5-21 is a child that would need to be studied further regarding (a)typical language development, considering her very low performance in both languages (see Öztekin, 2019).

The other 5-year-old who scored extremely low in Turkish comprehension was BiTur5-24 (MAIN1 = 1 point; MAIN2 = 0 points). Similarly to the previous child, BiTur5-24 had low Turkish vocabulary scores (CLT comprehension: 38/60, CLT production 14/60). She was born in Sweden and both parents also spoke Kurdish in addition to Turkish and Swedish. The parents (one Sweden-born and one Turkey-born) had both grown up in Sweden, and both had ISCED education level 4, but presumably a low level of literacy in Turkish. No storytelling activities in the home were reported for either language. Parental information about literacy activities with the child was contradictory ('no exposure to Turkish books' vs. 'occasional book reading in Turkish'). The parents reported that the child had 60% Swedish input during the day and that they did not focus on the improvement of the child's Turkish. BiTur5-24's Swedish comprehension scores were close to the means for her age group (MAIN1 = 8 points, MAIN2 = 6 points).

Another very low-scoring child in Turkish was 6-year-old BiTur6-03 (MAIN1 = 1 point; MAIN2 = 0 points, cf. age group means $M_{MAIN1} = 7.8$, $M_{MAIN2} = 6.3$). In Swedish though, the child received the maximum score for MAIN1 (10 points) and had a close to average score on MAIN2 (5 points). What else do we know of this child? BiTur6-03 had a very low Turkish vocabulary production score (CLT score: 12/60, cf. age group mean $M_{prod} = 39.5$). Her parents were Sweden-born and spoke both Turkish and Swedish with each other and the child. The child spoke mostly Swedish in the home. She was evaluated by her parents as more proficient in Swedish compared to Turkish, and to 'know very many words in Swedish'. Her daily input was 80% Swedish. The parents, both highly educated (ISCED level 7), reported regular book reading and storytelling in Swedish, but no such activities in Turkish. BiTur6-03 is thus a child with very low exposure to the heritage language. Table 4 illustrates her answers for the Turkish Dog story (see Table 2 for the questions). The child often answered *bilmiyorum* 'I don't know'. In the video recordings, she did not look as if she did not understand the questions but she looked rather timid. The reason why she did not answer the questions might thus have been that she did not feel confident in speaking Turkish.

Table 4. Answers to the Turkish MAIN1 (Dog) comprehension questions by BiTur6-03, a low-scoring child

Question	**Child answer (BiTur6-03, 6;0)**	**Score**
D1	*oraya atlıyor* 'jumping there'	0
D2	*bilmiyorum* 'I don't know'	0
D3	–	
D4	*o zaman balonu alacaktı* 'then (he) was going to take the balloon'	1
D5	*bilmiyorum* 'I don't know'	0
D6	–	
D7	*o zaman balonu alacaktı sonra onu gördü ve yedi* 'then (he) was going to take the balloon and then (he) saw it and ate (it)'	0
D8	*bilmiyorum* 'I don't know'	0
D9	–	
D10	*evet, bilmiyorum* 'yes, I don't know'	0
Total		**1**

The one 7-year-old who scored extremely low on comprehension of Turkish was BiTur7-01 (MAIN1 = 1 points, MAIN2 = 0 points). He had also some of the lowest scores on the Turkish vocabulary test in his age group (CLT comprehension: 40/60, production: 13/60). BiTur7-01 had close to average Swedish MAIN comprehension scores and above average Swedish vocabulary scores. This child lived with a single mother who was a native Swedish speaker, and received 80% Swedish input during the day. No Turkish book reading or storytelling activities occurred in the home. Again, this is a child with very little exposure to Turkish.

Summing up, children with exceptionally low *Turkish* MAIN comprehension scores for their age also performed very poorly on the Turkish vocabulary tasks. The questionnaire data indicate that these children had limited exposure to Turkish and few, if any, literacy-related or storytelling activities in Turkish. Sometimes, a third language was spoken in the home (Kurdish).[15] Alternatively, the parents spoke mainly Swedish with each other and with the child. Several of these children had one native Swedish or Sweden-born parent. These children might be on their way towards receptive bilingualism only, and not speak Turkish in the future.

Let us now move on to children who scored exceptionally low on the *Swedish* MAIN comprehension. These were mainly 4-year-olds, all with scores of 0–3 for both Swedish MAIN1 and MAIN2. These children also scored very low on the Swedish receptive and expressive vocabulary tasks. So what are these children's background characteristics?

BiTur4-23 was born in Turkey and moved to Sweden only 2.5 years before participating in the study. He attended a preschool with many other Turkish-speaking children and staff. His Turkish scores corresponded to the age group mean. At home, only Turkish was spoken by the parents, the child and his siblings. Turkish books were read and Turkish stories were told to the child on a daily basis by his highly educated parents (ISCED levels 7 and 6). No storytelling occurred in Swedish. Considering the child's late onset of Swedish and limited exposure, his low Swedish scores are not surprising.

Another very low-scoring child on Swedish MAIN was BiTur4-21. (Her Turkish scores were also low.) This child was born in Sweden and had started day-care at age two. However, the child had taken a lot of time off from preschool, for unclear reasons. Some of her preschool staff also spoke Turkish. Parents reported that they always spoke Turkish at home and because of this, they were worried about the child's Swedish. Both parents were Turkey-born, but one had grown up in Sweden. One parent could speak Kurdish but did not do so with the other parent or the children. Education levels were average (ISCED levels 4 and 3). Parents reported that no Turkish books were read to the child, and Swedish books only rarely. Storytelling sometimes occurred in Turkish, but never in Swedish. Based on this information, exposure to Swedish seems to have been limited, though somewhat confusingly, the child was reported to currently receive '80% daily input' in Swedish.

Another four-year-old, BiTur4-16, also had very low Swedish MAIN comprehension scores (but high Turkish scores). He was born in Sweden and had started preschool at age two, where he was exposed to Swedish and some Turkish. At home, the parents only spoke Turkish. Parental education was high. Book reading and storytelling activities were carried out with the child nearly every day in Turkish,

15. Note that low comprehension scores cannot be attributed to exposure to a third language per se, since many children in our sample who are exposed to three languages perform well. Rather, low narrative comprehension may be linked to limited language exposure, including limited narrative and literacy activities (see also Section 5).

but none in Swedish. BiTur4-16 had a late onset of speech and reportedly felt uneasy when choosing between Turkish and Swedish.

The last child worth mentioning, BiTur4-29, was born in Sweden and also started preschool at age two. Only Turkish was spoken in the home by the Turkey-born parents and they reported that the child had '95% input' in Turkish during the day since the child attended preschool only 25 hours per week. Parental education levels were average. Occasional literacy-related activities with the child were reported for both languages. The parents evaluated the child's Swedish ability as 'poor'. BiTur4-29's Swedish answers are exemplified in Table 5.

Table 5. Answers to the Swedish MAIN1 (Dog) comprehension questions, by BiTur4-29, a low-scoring child

Question	Child answers (BiTur4-29, 4;2)	Score
D1	*sen husmusen där gå in, sen hunden komma sen kryper in husmusen* 'then the house-mouse there goes in and then the dog come then crawls in the house-mouse'	0
D2	*nej [!] sa [?]* 'no [!] said'	0
D3	*jag in dä(r) [?], jag in dä(r)* [Dog's direct speech] 'I in there, I in there'	0
D4	*sen ta ballongen* 'then take the balloon'	1
D5	*yeah!*	0
D6	–	0
D7	*sen äta upp de korven* 'then eat the sausage'	1
D8	*nej* [Boy's direct speech] 'no'	0
D9	*komma, äta upp hunden* 'come, eat up the dog'	0
D10	*ja* 'yes'	0
Total		**2**

BiTur4-29 answers were very lively, showing that he put himself into the story character's shoes, using direct speech. However, his Swedish vocabulary and grammar were undeveloped for his age, with many omissions of function words (copula verbs, pronominal subjects), omission of inflections (e.g. tense) and word order difficulties. The low Swedish MAIN comprehension score may have been due to poor comprehension and/or insufficient language skills in answering the questions. (The child's Turkish MAIN1 score was above average.)

Only one older child performed exceptionally low on Swedish MAIN comprehension, namely 7-year-old BiTur7-26, who scored below 50% on MAIN2 (4.82 points, cf. age group mean $M_{MAIN2} = 8.3$). His Swedish vocabulary scores were also low (CLT comprehension: 42/60, production: 19/60; cf. age group means $M_{compr} = 56.8$, $M_{prod} = 42.4$). (His Turkish scores were close to average.) This child had not started to hear Swedish until age 5 when his family moved to Sweden, which means that he had only been exposed to Swedish for two years. Kurdish was the L1 of both parents, alongside Turkish. Parental education was very low (ISCED level 1), which may indicate low literacy. Very little book reading or storytelling activities were reported.

These snapshots of children with exceptionally low *Swedish* MAIN comprehension scores show that they also had low Swedish vocabulary scores and relatively little exposure to Swedish, due to late onset of Swedish and/or limited preschool attendance.

4.2.2 *Exceptionally high-scoring children*

Let us now consider the backgrounds of some exceptionally high-performing young children. Two four-year-olds stood out with their high *Turkish* comprehension scores. One of them was BiTur4-03 (MAIN1 = 10 points, MAIN2 = 6 points; cf. the means for his age group M_{MAIN1} = 5.8, M_{MAIN2} = 4.2). The parents spoke only Turkish with each other and Turkish and Swedish with the child. The child spoke mostly Swedish with the sibling. It was reported that the child heard 60% Turkish during the day and the parents evaluated the child's Turkish as 'very good'. One parent had grown up in Turkey, the other had been born and grown up in Sweden. Parental education was average (ISCED level 3). The child did many literacy- and language-related activities in Turkish and Swedish, including daily storytelling and shared book reading. The child's comprehension of Swedish MAIN was also above average for his age group (MAIN1 = 8, MAIN2 = 5). Table 6 shows the answers of this high-scoring 4-year-old for the Turkish Dog story.

Table 6. Answers to the Turkish MAIN1 (Dog) comprehension questions by BiTur4-03, a high-scoring 4-year-old child

Question	Child answers (BiTur4-03, 4;7)	Score
D1	*çünkü fareyi yakalamak için* 'because to catch the mouse'	1
D2	*köpek kötü hissederdi* 'the dog would feel bad'	1
D3	*ağaca çarptı* '(dog) hit the tree'	1
D4	*balonu almak için* 'to get the balloon'	1
D5	*iyi* 'good'	1
D6	*çünkü topunu yakaladığı için* 'because (he) could get his ball'	1
D7	*çünkü acıktı* 'because (he) was hungry'	1
D8	*kötü hissederdi* '(he) would feel bad'	1
D9	*çünkü köpeğe kızdı* 'because (he) is angry with the dog'	1
D10	*evet ama sosislerini yemezse arkadaş olur* 'yes but only if (the dog) does not eat (his) sausages'	1
Total		**10**

The other 4-year-old (BiTur4-15) with very high Turkish scores had two Turkey-born parents who spoke only Turkish in the home. She scored maximum points in comprehension of Turkish MAIN1 (10 points) and above average on MAIN2 (7 points). The child spoke mostly Turkish with her parents but also Swedish with her sibling. Parental education was average (ISCED level 3). The child was exposed to Turkish books and the parents reported frequent storytelling and book reading for both languages. The parents evaluated the child's Turkish as 'very good' and Swedish as 'poor'. However, her Swedish comprehension scores were above or around average (MAIN1 = 8 points, MAIN2 = 4 points).

Let us now move on to children with exceptionally high *Swedish* MAIN scores. Since Swedish comprehension was generally quite high among the older children, exceptionally high-scoring children only stand out in the youngest group, the 4-year-olds. BiTur4-05 not only had very high Swedish comprehension scores for her age (MAIN1 = 8 points, MAIN2 = 8 points) but also exceptionally high Swedish CLT vocabulary scores (60/60 for comprehension, 47/60 for production). This child had one native-Swedish parent, and both Turkish and Swedish were spoken in the home. The child had started attending Swedish preschool at 1;4 in an affluent district and did so 48 hours a week, the highest attendance of all children in our sample. BiTur4-05 was described as a 'very fast developer' and reported to receive 80% Swedish input during the day. Her parents had the highest education in the study (ISCED levels 7 and 8). The child did a lot of activities (book reading, watching TV, singing, telling stories) in both Turkish and Swedish almost every day. Her Turkish MAIN comprehension scores (MAIN1 = 8 points, MAIN2 = 8 points) were also far above the mean for her age group. In Table 7, the Swedish answers of this child to the comprehension questions for the Dog story are shown.

Another four-year-old who scored very high on the comprehension of Swedish MAIN was BiTur4-30 (MAIN1 = 9 points, MAIN2 = 6 points). Surprisingly, only Turkish was spoken in the home and the parents had moved to Sweden less than ten years ago. The child started hearing Swedish at age one when he began daycare. Parents reported that the child developed very fast in both languages and spoke better than his peers. Information on storytelling and literacy-related activities was missing. The child performed below average on the comprehension of Turkish (MAIN1 = 4 points, MAIN2 = 4 points).

What these case studies of individual children tell us is that very good or very poor comprehension results on MAIN often go together with very high or very low vocabulary results in that language, and with characteristics in the children's language backgrounds (in terms of amount of exposure to the languages), and potentially also with the cultural habits of families concerning joint book reading and storytelling in the home (though these would need to be explored in greater

Table 7. Answers to the Swedish MAIN1 (Dog) comprehension questions by BiTur4-05, a high-scoring 4-year-old child

Question	Child answers (BiTur4-05, 4;6)	Score
D1	*för att den vill ehm tag i musen* 'because he wants to get hold of the mouse'	1
D2	*inge(t) bra* 'not good'	1
D3	*för att det där lilla hålet var för litet* 'because that little hole was too small'	1
D4	*för att ta sin ballong* 'to take his balloon'	1
D5	*jättebra* 'very good'	1
D6	*för att han fick tag i sin ballong* 'because he got hold of his balloon'	1
D7	Experimenter forgot to ask this question	0.81*
D8	*inge(t) bra* 'not good'	1
D9	*för att hunden åt upp korven* 'because the dog ate the sausage'	1
D10	*för att (.) han blev kompis när han slog sig i huvet* 'because he became friends when he hit his head'	0
Total		**8.81**

Note. * As this was a case of experimenter error, the mean score for this question for Swedish MAIN1 (Dog) for 4-year-olds ($M = 0.81$) was substituted.

depth than is possible here). In the following section, the relation between MAIN comprehension scores and vocabulary knowledge is investigated statistically for all children.

4.3 Influence of vocabulary knowledge on MAIN comprehension scores

We explored the effects of vocabulary comprehension and production on the MAIN comprehension scores, while controlling for age (in months) and the language of the first testing. Table 8 summarises the 4 linear regression models, one for each of the two tasks (Cat/Dog, Baby Birds/Baby Goats) in each language.

For both Turkish narrative tasks, Cat/Dog and Baby Birds/Baby Goats, there were significant effects of age in months and *vocabulary production* scores (CLT) on the narrative comprehension scores. The effects of vocabulary comprehension and the language of the first testing were not significant, what mattered was expressive vocabulary. The linear regression models explained a relatively large part of the variance in the children's scores (Cat/Dog: 45.6%, Baby Birds/Baby Goats: 45.5%). The models thus confirm the picture painted by the snapshots of individual children in Section 4.2.

Table 8. Summary of linear regression models for the comprehension scores of the narrative tasks in Turkish and Swedish

	Tur Cat/Dog[a] (*N* = 100)		**Tur BB/BG**[b] (*N* = 99)		**Swe Cat/Dog**[c] (*N* = 99)		**Swe BB/BG**[d] (*N* = 99)	
	β	*SE*	*β*	*SE*	*β*	*SE*	*β*	*SE*
Intercept	−2.87	1.99	−4.84*	2.19	−0.88	1.42	−5.73***	1.26
Age (months)	0.04**	0.01	0.05**	0.02	0.03*	0.02	0.08***	0.01
Vocab – comp	0.08	0.05	0.07	0.05	0.09*	0.05	0.06	0.04
Vocab – prod	0.08**	0.03	0.10**	0.03	0.03	0.03	0.08**	0.03
Test1	−0.67	0.36	−0.55	0.40	0.92**	0.34	0.43	0.30
R^2 (adjusted)	.456		.455		.413		.675	

Note. *** = $p < .001$; ** = $p < .01$; * = $p < .05$.
BB = Baby Birds, BG = Baby Goats, Vocab – comp = receptive CLT score in the same language as the narrative task, Vocab – prod = expressive CLT score in the same language as the narrative task, Test1 = language of the first testing; the model shows the effect when the language of the first testing is Turkish.
a. $F(4, 95) = 21.75$, $p < .001$, b. $F(4, 94) = 21.41$, $p < .001$, c. $F(4, 94) = 18.23$, $p < .001$, d. $F(4, 94) = 51.98$, $p < .001$.

In Swedish, similar effects were found: For Swedish Baby Birds/Baby Goats, there were significant effects of age (in months) and vocabulary production scores on the narrative comprehension scores. In fact, the model explained an even larger part of the variance (67.5%) in MAIN comprehension scores than for Turkish. This confirms what we found in the case studies of individual children.

However, the results for Swedish Cat/Dog were different. Vocabulary production scores had no significant effect, but in addition to age, children scored significantly better on comprehension of Swedish Cat/Dog if they had higher scores on the comprehension parts of the CLT and when Turkish was the language of the first testing ($M = 8.2$) than when Swedish was tested first ($M = 7.3$).[16] We are

16. Why is the language of the first testing an additional (significant) predictor of the MAIN comprehension score only for Cat/Dog in Swedish? Here is a (speculative) explanation: Recall that Cat/Dog is the very first task in the Swedish test session. After some brief warming up chat, a stranger, the experimenter, asks the child to tell a story from pictures in an unfamiliar format and to answer comprehension questions. It is a new type of task for the child and it is also done in Swedish, which for at least 39% of the children is weaker than their Turkish, according to the parents. For those children that were tested in Swedish first (due to our counterbalancing), Swedish Cat/Dog is in fact the *very first task* they have to do. For the three subsequent MAIN tasks in Swedish and Turkish, they will already be familiar with the format. In Turkish, even when Turkish was tested first, Cat/Dog was not the very first task for the child but only the first narrative task. In Turkish, after the initial warming up chat, the child first did a non-word repetition parrot game with the experimenter before Cat/Dog was administered. This might be why the language of the first testing is an additional predictor of the comprehension score for Cat/Dog only in Swedish.

currently unsure why the Swedish Cat/Dog scores at group level are not influenced by expressive vocabulary knowledge as the other narrative tasks. We return to the vocabulary findings in the discussion.

4.4 Performance on individual comprehension questions

Recall that we not only found effects of age, but also effects of task, with significantly higher overall comprehension scores for Cat/Dog (MAIN1) than for Baby Birds/Baby Goats (MAIN2) in both languages (Section 4.1). We therefore decided to explore this task effect further, by taking a closer look at the individual questions in Cat/Dog vs Baby Birds/Baby Goats. All four stories are depicted in Figures 8 and 9.

Figure 8. Small-scale copies of MAIN Cat (left) and Dog (right) (Gagarina et al., 2012) (reproduced with permission from the publisher)

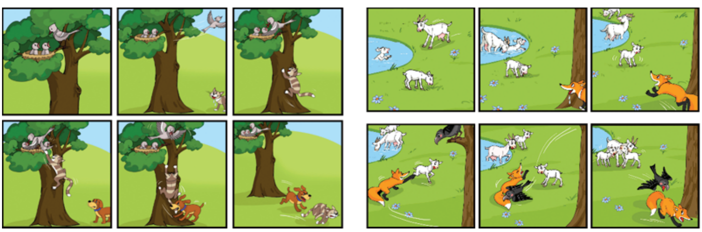

Figure 9. Small-scale copies of MAIN Baby Birds (left) and Baby Goats (right) (Gagarina et al., 2012) (reproduced with permission from the publisher)

4.4.1 *Response accuracies for individual comprehension questions*

Let us now investigate performance on individual comprehension questions, starting with the Cat/Dog task. Table 9 shows the numbers and percentages of correct answers for each question. (For the questions, the reader is referred back to Table 2.) The answer patterns in the two languages are strikingly similar. In both Turkish and Swedish, most questions were answered correctly by the majority of children (with accuracies above or around 60%, often above 70%). The most accurately answered question was D5 in both languages (90% accuracy), which queried an internal state as reaction (*How does the boy feel?*), closely followed by D7 (87%), which queried the goal in Episode 3 (*Why does the cat/dog grab the fish/sausages?*). By far the lowest response accuracy in both languages (34%, 35%) was found for the final question in Cat/Dog, D10 (*Will the boy be friends with the cat/dog? Why?*).

Table 9. Cat/Dog (MAIN1), number and percentage (%) of correct answers for the ten comprehension questions in Turkish and Swedish. All children combined

Question	Turkish (*N* = 100)	Swedish (*N* = 99)
D1. Episode 1 Goal	80 (80%)	84 (85%)
D2. Episode 1 Internal state	83 (83%)	77 (78%)
D3. Episode 1 Internal state rationale	61 (61%)	71 (72%)
D4. Episode 2 Goal	73 (73%)	81 (82%)
D5. Episode 2 Internal state	90 (90%)	89 (90%)
D6. Episode 2 Internal state rationale	86 (86%)	84 (85%)
D7. Episode 3 Goal	87 (87%)	86 (87%)
D8. Episode 3 Internal state Theory of Mind	64 (64%)	77 (78%)
D9. Episode 3 Internal state rationale (Theory of Mind)	56 (56%)	69 (70%)
D10. Overall plotline (Theory of Mind)	35 (34%)	36 (35%)

Accuracy on questions targeting the comprehension of goals (D1, D4, D7) and internal states (D2, D5) including theory of mind questions (D8, D9) was high for both Turkish and Swedish. The children's overall comprehension of Cat/Dog was thus good, except for the final question (D10).

Let us now compare this to Baby Birds/Baby Goats. Table 10 shows the number and percentage of correct answers for each comprehension question.

Just as for Cat/Dog, the answer patterns in Turkish and Swedish are strikingly similar. However, accuracy was generally somewhat lower in Baby Birds/Baby Goats than in Cat/Dog. By contrast, comprehension of the final question (D10) was better.

D5 (*How does the cat/fox feel?*) had the highest accuracy rate (81%, 86%) in both languages, just as it did in Cat/Dog. Equally high percentages correct (85%) were found in both languages for D4, which queried the goal in Episode 2 (*Why does the cat climb the tree/why does the fox jump forward?*).[17]

In both languages, the lowest accuracy was found for Baby Birds/Baby Goats D9 (27%, 35%), which queried the rationale for a character's internal state at the end of Episode 3. D8, a 'what if' preamble to D9, was not often answered correctly

17. For both Baby Birds/Baby Goats and Cat/Dog, questions targeting goals (from two different episodes) thus yielded some of the highest response accuracies.

Table 10. Baby Birds/Baby Goats (MAIN2), number and percentage (%) of correct answers on the ten comprehension questions in Turkish and Swedish. All children combined

Question	Turkish ($N = 99$)	Swedish ($N = 99$)
D1. Episode 1 Goal	75 (76%)	66 (67%)
D2. Episode 1 Internal state	71 (72%)	70 (71%)
D3. Episode 1 Internal state rationale	46 (47%)	39 (40%)
D4. Episode 2 Goal	84 (85%)	84 (85%)
D5. Episode 2 Internal state	80 (81%)	85 (86%)
D6. Episode 2 Internal state rationale	66 (67%)	66 (67%)
D7. Episode 3 Goal	42 (43%)	65 (66%)
D8. Episode 3 Internal state Theory of Mind	50 (51%)	51 (53%)
D9. Episode 3 Internal state rationale (Theory of Mind)	26 (27%)	34 (35%)
D10. Overall plotline (Theory of Mind)	55 (56%)	52 (53%)

either; and again this held for both languages (51%, 53% accuracy). D8: *Imagine that the dog/crow sees the birds/goats now. How would the dog/crow feel? >> D9: Why would the dog/crow feel* [answer D8]*?*

D3, which queried the rationale for a character's internal state in Episode 1 (*Why do the baby birds feel/Why does the baby goat feel …?*), also had low accuracies in both languages (47%, 40%). Possible reasons for the low performance on these particular comprehension questions are discussed below.

Summing up, just as for Cat/Dog, the answer patterns for individual questions in Baby Birds/Baby Goats were extremely similar in the two languages.[18] However, the answer patterns differed for the two tasks (Cat/Dog vs Baby Birds/Baby Goats), with lower response accuracies on most questions in Baby Birds/Baby Goats compared to Cat/Dog, in line with the task effect found for the overall comprehension scores (Section 4.1).

18. However, one difference deserves attention. The accuracy for Baby Birds/Baby Goats D7 (*Why does the dog/crow grab the cat's/fox's tail?*) was considerably lower in Turkish (43%) than in Swedish (66%). When exploring the scores of individual children, we noticed that zero scores in Turkish on D7 mainly came from the 4- and 5-year-olds, but the 6- and 7-year-olds also performed worse in Turkish than in Swedish. D7 accuracy rates in Turkish were for the 4-years: 12%, 5-years: 36%, 6-years: 50%, 7-years: 67%, vs. in Swedish, 4-years: 39%, 5-years: 77%, 6-years: 71%, 7-years: 81% correct. The answers by the 4- and 5-year-olds in Turkish were incorrect for a number of different reasons. Some responses were not an answer to the question, while others were too vague or too general to merit a point according to the scoring protocol, such as *çünkü tilki çok kötu bir şey yaptı* 'because the fox did something very bad' or *çünkü o onu alacaktı* 'because it is going to take it'.

Why did the children answer the majority of questions less accurately in Baby Birds/Baby Goats than in Cat/Dog, even though they were all administered in exactly the same mode? And why did the children perform so poorly on the last question (D10) in Cat/Dog? Here we need to take a closer look at the individual comprehension questions, what they imply, what inferences the child has to make to score a point, and what a child can read off the pictures. This is done in the following section.

4.4.2 *Exploring individual comprehension questions further*

Let us now investigate the questions with particularly low response accuracies in relation to what is depicted in the stimulus pictures and what kind of inference is necessary to answer the question correctly. We also take a cursory look at how the different age groups perform on these particular comprehension questions concerning accuracy and types of incorrect answers, in order to discover patterns and development with age, if any.

Let us begin with Cat/Dog D10, the only Cat/Dog question with very low accuracy (lower than for all the other questions in Cat/Dog, and also much lower than D10 in Baby Birds/Baby Goats). D10 asks whether the boy would become friends with the cat/dog and why/why not. The expected answer is 'No, they won't be friends because the cat/dog ate/stole the boy's fish/sausages'.[19]

Here an interesting difference between younger and older children emerges: Whereas very few 4- and 5-year-olds in our sample answered this question correctly (4 years: Turkish 8%, Swedish 17%; 5 years: Turkish 18%, Swedish 23%), the 6- and 7-year-olds answered D10 correctly half of the time (6 years: Turkish 50%, Swedish 50%; 7 years: Turkish 57%, Swedish 50%). Of the 4- and 5-year-olds, only few could answer the why-question in any comprehensible way at all; quite a few said that they didn't know. Some children approached the situation differently and answered that the boy and the dog/cat would be friends because the dog/cat was very cute or fun to be with (e.g. Tur. *evet, çünkü çok iyi bir köpek* 'Yes, because it is a very good dog'; *olur, çünkü mutlu bi kedi bu* 'Would be [friends], because this is a very happy cat'; *olur, çünkü tatlı olduğunu düşündü* 'Would be [friends], because he thought that it was so cute'; Swe. *ja, den är snäll och rolig* 'Yes, [because] it is kind and fun'). These children may be familiar with cats and dogs as pets and approach them with positive emotions. (Such answers are scored as incorrect on MAIN.) In the incorrect answers of the older children, particularly the 7-year-olds, logical reasoning explanations predominated: For instance, children said that the boy and

19. Answers of the type 'Yes, they will be friends because the boy will forgive the cat/dog' are accepted as well.

the cat weren't friends because they came from different directions in the pictures (e.g. Swe. *därför han kom från nånstans och katten kom från annanstans* 'because he came from somewhere and the cat came from somewhere else'). Older children also frequently answered that the boy would be friends with the dog/cat because it is in fact the boy's pet (e.g. Swe. *jag tror det är hans katt* 'I think it is his cat'; Tur. *evet, çünkü o onun köpeği* 'Yes, because it is his dog'; *olur, belki kendi köpeğidir, çünkü tasması var* 'Would be [friends], because maybe it is his dog, because it has a leash'). While such pet answers are not directly based on the events shown in the picture sequence and are not the expected answers, they are based on accurate world knowledge: In the real world, dogs and cats often do belong to people, and if you have them as a pet, you tend to be friends with them and have them close by.

For Cat/Dog D10 then, the very low response accuracy for all children combined (34%, 35%) masks an age development, which would deserve a more detailed investigation. The older children's logical but 'incorrect' responses moreover suggest that scoring a point on D10 can be difficult when the child 'goes outside' the plotline shown in the MAIN pictures.[20]

Let us now move on to the – relatively many – questions that were answered less accurately in Baby Birds/Baby Goats than in Cat/Dog (recall Tables 9–10). Table 11 discusses these questions contrastively.

The contrastive analysis in Table 11 pinpoints certain differences between Cat/Dog and Baby Birds/Baby Goats: differences in the way facial and bodily reactions of characters are depicted (or not depicted), differences in whether an event leading up to an internal state is depicted (or not), differences in human vs non-human agents, and differences in the prototypicality of certain behaviours in real life. We suggest that these differences may explain the children's lower performance on D3, D6, D7, D8 and D9 in Baby Birds/Baby Goats, since for each of these questions, the child has to 'do more work' in inferring the story character's internal state in Baby Birds/Baby Goats than in Cat/Dog.

20. Lindgren and Bohnacker (this volume) also investigate this issue for German/Swedish bilinguals.

Table 11. MAIN comprehension questions that were answered substantially less accurately in Baby Birds/Baby Goats in both languages, and possible explanations for the differences

Question	Cat/Dog	Baby Birds/Baby Goats
D2 / D3	*How does the cat/dog feel here?* [picture 3] >> *Why do you think that the cat/dog is feeling bad (angry/hurt)?*	*How do the baby birds/ How does the baby goat feel here?* [picture 1] >> *Why do you think that the baby birds are feeling bad (hungry)/ the baby goat is feeling bad (scared)?*
	The queried internal state is a reaction to an event. The picture shows that the cat is in the bush, and the dog hits the tree. We can see both their faces. The painful facial expressions and scrunched-up bodies are easily recognised as discomfort or pain. The event chain leading to the internal state ('angry/hurt') is explicitly shown in the preceding pictures, i.e. the reaction comes after the cat fell into the bush and after the dog hit its head on the tree. The resulting feeling is graphically shown in the whole body of the cat/dog.	The queried internal state is not a reaction to an event, but a trigger for the following event. In *Baby Birds*, the picture shows that the baby birds are in a nest in the tree, but the only facial expression we can see is that their mouths are open. From this open mouth cue the child has to infer a feeling of discomfort/hunger. The facial expression of the mother bird is neutral; the child has to infer that she flies away in order to bring food. In *Baby Goats*, the baby goat is in the water and the mother goat is running towards it. Both have worried looks on their faces. The child has to infer that the baby goat is scared as a result of being in the water. No event chain that leads to the internal state is depicted, in other words, no activity is shown that leads to the goat being in the water.[21]

21. Note the difference between Baby Birds and Baby Goats here. It may be more difficult for a child to infer from Picture 1 that the baby birds are feeling hungry/bad, as the open mouth cue may also indicate that the birds are (happily) singing or chirping. In Baby Goats, inferring that the baby goat is feeling scared/bad in the water might be easier because the goat has an unhappy, worried look on its face. The mother goat, who is running towards it, has also a worried facial expression. These multiple visual cues are clearer in Baby Goats than those provided in Baby Birds. However, neither depiction is as easily recognised as the cat's/dog's discomfort/pain in Cat/Dog. We believe that these pictorial differences may be the reason for the lower performance on (Turkish) Baby Birds than on Baby Goats (Section 4.1, Figure 3). When the answers for each question were examined separately for Baby Birds vs Baby Goats, accuracy was much lower in Baby Birds precisely for D2 and D3 (D2 Baby Birds = 56% correct vs Baby Goats = 89%; D3 Baby Birds = 35% correct vs Baby Goats = 60%). The lower accuracy on D2 predicts lower accuracy in D3, since D3 is only asked if D2 has been answered correctly. Incorrect answers to D2 in Baby Birds often took the format *anne diyorlardı* 'they are saying mom!', *şarkı söylüyorlardı* 'they were singing', or *iyi hissederlerdi çünkü anne orda* 'they would feel good because the mother is there.'

Table 11. (*continued*)

Question	Cat/Dog	Baby Birds/Baby Goats
D6	*Why do you think that the boy is feeling good/happy/content?* [picture 6]	*Why do you think that the cat/fox is feeling bad/angry/sad?* [pictures 5–6]
	The character is a person, a boy, and his smiling facial expression can easily be recognised as 'happy'. The boy gets his balloon/ball back, which is easy for children to relate to since they play with such objects as toys and are likely to be familiar with losing/finding them.	The characters are animals. It may not be as easy to read the face of these animals. In Baby Birds, the child cannot see the face of the cat that is hurt/angry/sad in picture 5.
D7	*Why does the cat/dog grab the fish/sausages?* [picture 5]	*Why does the dog/crow bite the cat's/fox's tail?* [picture 5]
	The cat/dog eats the fish/sausages that the boy left on the grass. The correct answer to D7 is as simple as 'because the cat/dog wanted to eat them/because the cat/dog was hungry'. The goal implies a basic, internal physiological state (hunger). The child can easily understand that the animals are hungry and therefore eat the fish/sausages.	The dog/crow bites the tail of the cat/fox that is trying to catch/eat the bird/goat, which is not a frequent daily occurrence. In order to answer D7 correctly, the child has to infer and understand the deeper inner motivation of the dog/crow, i.e. wanting to rescue/help the baby bird/goat.
D8 / D9	*Imagine that the boy sees the cat/dog. How would the boy feel?* [picture 6] >> *Why do you think that the boy feels bad/angry?*	*Imagine that the dog/crow sees the birds/goats. How would the dog/crow feel?* [picture 6] >> *Why do you think that the bird/dog feels happy/good/content/like a hero?*
	The correct answer to D8 is 'bad/angry.' The one who sees the cat/dog is a person, the boy. In order to answer D8 and D9 correctly, the child has to infer a negative emotion as reaction in the boy, because the cat/dog stole his food. This should be easy for children to relate to everyday life. For such an inference, the child does not need to keep the entire plotline in mind. Assigning emotions to a person may also be easier than assigning them to an animal.	Correct answers to D8 are 'good/happy/relieved/proud/like a hero'. The one who sees is an animal, the dog/crow. In order to answer D8 and D9 correctly, the child has to attribute human feelings to an animal as well as understand the inner motivation of the dog/crow, which is to rescue the baby bird/goat. Here, the child needs to keep in mind the entire plotline (pictures 1–6). Only then can the child arrive at the correct positive internal state answer to D8 ('happy') and the correct answer to D9 ('because the bird/goat was saved/because the attacker was driven away').

Finally, Table 12 shows how our participants performed on D3, D6, D7, D8 and D9 in Baby Birds/Baby Goats, broken down by language and age group. For all questions, response accuracy increases with age, though it varies how steeply accuracy increases at which age.

Table 12. Baby Birds/Baby Goats comprehension questions with low accuracy rates, number and percentage (%) of correct answers, by age group in Turkish and Swedish

Question		4 years	5 years	6 years	7 years
D3	Turkish	6 (25%)	12 (57%)	14 (54%)	14 (50%)
	Swedish	5 (22%)	3 (14%)	13 (50%)	18 (64%)
D6	Turkish	11 (46%)	16 (76%)	17 (65%)	22 (79%)
	Swedish	9 (39%)	13 (59%)	18 (69%)	26 (93%)
D7	Turkish	3 (13%)	8 (38%)	13 (66%)	18 (44%)
	Swedish	9 (39%)	17 (77%)	17 (65%)	22 (79%)
D8	Turkish	11 (47%)	9 (43%)	13 (50%)	17 (61%)
	Swedish	4 (17%)	10 (45%)	16 (62%)	21 (75%)
D9	Turkish	3 (13%)	5 (24%)	6 (23%)	12 (43%)
	Swedish	1 (4%)	7 (32%)	10 (38%)	16 (57%)

Consider for instance D6 (*Why do you think that the cat/fox is feeling bad/angry/sad?*), where percentages correct increase for almost every age group. When we analysed the *types* of answers to D6, incorrect responses by the youngest children were often no answer at all, or they were incomprehensible (either lengthy rambling or vague, rudimentary formulations (e.g. Swe. *äta* 'eat')). A particular, frequent type of incorrect answer to D6 was: 'the cat feels bad because the dog wants to eat the cat' (e.g. Swe. *därför hunden vill äta han hela* 'because the dog wants to eat him whole'; Tur. *çünkü kediyi yemek istiyor* 'because (it) wants to eat the cat'). Such answers indicate that the child has not understood the role of the dog/crow as the intervening helper or rescuer in the story. Some of the youngest children also incorrectly answered that the dog/crow wants to eat the baby bird/baby goat himself (e.g. Tur. *çünkü onu yemek istiyo(r)* 'because it wants to eat it' (child points to baby goat); *çünkü kuzuyu yemek için* 'because to eat the lamb [= baby goat]'). Again, the child has not understood the role of the dog/crow as it was intended. Such answers did not occur in the older children.[22]

Finally, let us consider the answers to D8 and D9, asked in connection with the last picture (D8: *Imagine that the dog/crow sees the birds/goats. How would the dog/*

22. For similar age-related findings in German-Swedish bilinguals, see Lindgren & Bohnacker (this volume).

crow feel? >> D9: *Why would the bird feel …?*). The expected answer here is a positive emotion (e.g. *good, happy, relieved, pleased, proud, like a hero, glad* – because the birds/goats are safe now or because s/he helped/rescued the baby bird/goat). As Table 12 shows, accuracy rates were lower for the younger children but increased with age. Incorrect responses by the youngest children to D8/D9 were often no answer at all or 'I don't know'. Another, interesting type of incorrect answer was that the dog/crow feels 'angry/cross/not good' (e.g. Tur. *köpek kötü hissederdi* 'the dog would feel bad', *mutlu değil* 'not happy', *kızgın* 'angry', *hiç iyi değil* 'not good at all', *üzgün* 'sad'; Swe. *arg* 'angry', *jättearg* 'very angry'). Such negative emotional state terms as answers were very frequent in the 4- and 5-year-olds. Answers expressing negative emotions ('angry') indicate that the child does not pay attention to the hypothetical preamble (*Imagine that the dog* [point to dog] *sees the birds* [point to birds]. *How would the dog feel?*). Instead, the child focuses on the dog's facial and bodily expression in picture 6 and infers an internal state from it (as has been pointed out by Bohnacker (2016) and Bohnacker & Lindgren (in press)). Moreover, the child ignores the previous plotline of the story, having forgotten or not understood that the dog/crow, upon seeing that the baby bird/goat is in danger, intervenes and rescues the little one from the clutches of the cat/fox. Negative emotions as incorrect answers to D8 were frequent in the younger children, but rarer in the older age groups (6 and 7 years).[23]

The above findings suggest that certain individual comprehension questions on MAIN are indeed harder for the children, as reflected in the lower response accuracies. Our qualitative analysis, which contrasted the pictorial stimuli in the different MAIN stories and the mental work required by the child to infer the internal state of a story character on a particular comprehension question, moreover suggests that several seemingly identical comprehension questions are harder in Baby Birds/Baby Goats than in Cat/Dog.

5. Discussion and conclusion

In this chapter, we have investigated story comprehension in 100 bilingual Turkish-Swedish children in both languages, using comparable tasks (MAIN picture sequences and inferential *how* and *why* questions). The cross-sectional setup included children aged 4;0–8;1.

Similarly to studies of other language combinations (e.g. Bohnacker, 2016; Kapalková et al., 2016; Bohnacker & Lindgren, in press; Fiani, Henry, & Prévost,

23. For similar findings concerning bilingual German-Swedish, see Lindgren & Bohnacker (this volume).

this volume; Kunnari & Välimaa, this volume; Lindgren & Bohnacker, this volume), there was no main effect of language; i.e. at group level the children scored similarly in Turkish and Swedish narrative comprehension.

Overall comprehension scores increased significantly with age in both languages, which we interpret to mean that children's story comprehension measurably improves with age. This overall age development parallels the results of earlier studies of smaller groups of bilingual English-Swedish, German-Swedish, Italian-English and Turkish-German children tested with MAIN (Bohnacker, 2016; Maviş et al., 2016; Roch et al., 2016; Lindgren, 2018), and is documented here for the first time for Turkish and Swedish on a larger scale. Our results are also congruent with large-scale studies of story comprehension in monolingual children of similar ages in other settings that employed different narrative materials (e.g. Hayward et al.'s 2009 cross-sectional study of monolingual English children age 4, 5 and 6 in Canada; Westerveld et al.'s 2012 longitudinal study of monolingual English children age 4–5 in New Zealand; Lepola et al.'s 2012 longitudinal study of monolingual Finnish children age 4, 5, and 6 in Finland).

The MAIN comprehension questions probe the child's ability to identify and encode causal connections in picture-based stories. To make these inferences, children need an understanding of the psychological and motivational causes of story characters' actions and to connect causes and consequences within an episode and between episodes (Lynch et al., 2008). The increase in comprehension scores in both Turkish and Swedish suggests that the children improve their inferencing abilities with age in both languages. Indeed, with increasing age and cognitive maturity, accumulating experience and background knowledge, children become more adept at inference generation and reasoning and can thus make more effective use of their limited attentional or working memory capacities in story comprehension tasks (cf. Trabasso et al., 1984; van den Broek et al., 2005).

Beside an overall improvement with age in both languages, the bilingual children's scores distributed somewhat differently – more scattered in Turkish and more clustered in Swedish, with a clearer increase between age groups in Swedish than in Turkish. More of the younger children (age 4–5) performed well in Turkish, whilst more of the older children (age 6–7) performed well in Swedish. In this cross-sectional study, story comprehension thus improved more visibly in the majority language Swedish than in Turkish. A possible explanation could be that the children were attending Swedish-medium (pre)schools for a major part of the day, where curriculum activities include shared book reading and storytelling. This may have boosted the children's Swedish narrative abilities and led to more uniform scores in Swedish at age 6–7. In the minority (and home) language Turkish, input and home-based activities that foster narrative abilities are likely to have been more varied from child to child.

In both languages, the older children in our sample (age 6–7) achieved relatively high story comprehension scores (age 6: Swe. $M_{MAIN1} = 8.6$, $M_{MAIN2} = 7.1$ and Tur. $M_{MAIN2} = 7.8$, $M_{MAIN2} = 6.3$; age 7: Swe. $M_{MAIN1} = 8.8$, $M_{MAIN\ 1} = 8.4$ and Tur. $M_{MAIN1} = 7.9$, $M_{MAIN2} = 7.2$, Figure 2), which suggests that they understood the pictorially presented stories quite well. As the MAIN comprehension questions probe inferential appreciation of the motivations and internal states of story characters, high scores indicate that the children were able to make such inferences. However, even though some children reached the maximum score (10/10), it was not the case that the majority of children performed at ceiling, not even in the oldest age group (age 7). This indicates that there are some aspects of inferential story understanding (as probed by MAIN) that are not fully mastered by age 6–7.

Indeed, response accuracies varied considerably between different types of comprehension questions, whereas overall response *patterns* were extremely similar across languages. This indicates that children's story comprehension performance is not only related to age, but also related to task, and in particular to individual comprehension questions. Individual comprehension questions in MAIN require rather different types of inferences. Here, the 'leg-up' a child gets from the stimulus pictures and the mental work required to make that inference varies greatly between questions. We found that questions were more often answered correctly when the motivation and internal states of a story character could be easily 'read off' the stimulus picture and/or matched frequent, prototypical behaviours of humans and animals. By contrast, response accuracy was often low when the inference necessitated that the child abstract away from the facial expressions and event(s) shown in one picture, and instead take several episodes or the entire plotline into account (as previously noted by Bohnacker, 2016; Bohnacker & Lindgren, in press).

Besides differences in the performance on *individual* comprehension questions, we also found a clear *overall effect of task*. Two of the four picture sequences, Cat and Dog (MAIN1), are parallel in plotline and story grammar, and the other two picture sequences, Baby Birds and Baby Goats (MAIN2), are also parallel. Comprehension scores were significantly lower for Baby Birds/Baby Goats (MAIN2) than for Cat/Dog (MAIN1) in *both* Turkish and Swedish. The task influenced comprehension more than the language of testing did.

In Baby Birds/Baby Goats, the children answered the majority of the questions less accurately than in Cat/Dog; only on one question (D10) did they perform worse in Cat/Dog. This indicates that comprehension of Baby Birds/Baby Goats, as measured by MAIN, is generally more difficult. With our qualitative analysis, we have tried to pinpoint why parallel and seemingly identical inferential questions are often more difficult to answer in Baby Birds/Baby Goats than in Cat/Dog (Section 4.4.2). The reason is that there are subtle differences in the way facial and bodily reactions of characters are depicted, differences in whether an event leading up to an internal state is depicted (or not), human vs. non-human agents, and differences in the prototypicality of certain behaviours. We suggest that these differences explain the children's lower performance on Baby Birds/Baby Goats than on Cat/Dog *across languages*.[24] This is an important finding that goes beyond previous work.

When the MAIN picture sequences and stories were originally designed, they were intended to be structurally parallel, and the MAIN tasks, including the comprehension questions, were designed to be equally difficult (Gagarina et al., 2015: 256). Baby Birds/Baby Goats were originally intended for telling, and Cat/Dog for retelling. Yet when MAIN was launched in 2012/2013, the developers' assumption of the comprehension tasks being equally difficult for all stories had not been tested empirically.

Our results show that comprehension of Baby Birds/Baby Goats is more difficult than Cat/Dog when the stories are administered in exactly the same mode. This is an important finding, and it puts earlier reports of MAIN comprehension data in a new light. Several studies (e.g. Maviş et al. 2016; Roch et al., 2016; Otwinowska et al., 2018) have assessed comprehension of Baby Birds/Baby Goats after telling, and comprehension of Cat/Dog after listening to a model story (i.e. where the child first listens to the story and then answers the comprehension questions, or, alternatively, listens to the story and then retells it before being asked the comprehension questions). Having found higher scores for Cat/Dog, these higher scores have been interpreted to mean that story comprehension is easier after listening. However, the present study has shown that comprehension of Cat/Dog is easier than Baby Birds/Baby Goats *even when* Cat/Dog is *not* administered in the listening or listen-and-retell mode. This indicates that irrespective of telling, model story or retelling, the comprehension questions in Cat/Dog are easier (as is also argued by Bohnacker & Lindgren, in press; Lindgren, 2018; Lindgren & Bohnacker, this volume). When assessing a child's story comprehension with MAIN, performance on Cat/Dog can therefore not be straightforwardly compared with Baby Birds/Baby Goats. Future researchers and clinicians should keep this in mind when examining bilingual children's narrative abilities, so as not to arrive at an incomplete or misleading picture of the child's story comprehension in the two languages.

Despite the aforementioned finding of a clear overall increase in comprehension scores with age, scores varied greatly between individual children. To explore

24. The lower comprehension scores in Baby Birds than in Baby Goats (found for Turkish, Figure 3, Section 4.1), which could be traced back to lower response accuracies on D2 and D3 in Baby Birds, may also be due to pictorial differences (see Table 11, 4.4.2). Lower comprehension scores on Baby Birds than Baby Goats have also been found by researchers of other language groups (Bohnacker & Lindgren, in press, for English/Swedish; Lindgren, 2018, 2019, for Swedish), but without offering an explanation.

this variation further, we tested the influence of children's vocabulary knowledge on MAIN comprehension for both languages (Section 4.3). The results showed that both age and expressive vocabulary (CLT production) were significant predictors of story comprehension for all Turkish tasks and for one of the Swedish tasks. Expressive vocabulary knowledge explained a relatively large part of the variance in the children's story comprehension scores. This means that good comprehension results on MAIN often go together with good vocabulary scores in that language, and poor comprehension results on MAIN often go together with poor vocabulary scores in that language. Again this is an important and new result, because narrative comprehension has not previously been systematically studied in relation to vocabulary knowledge in bilingual children. So what does this result mean? It must mean that story comprehension, as measured on MAIN, is dependent on lexical knowledge, particularly expressive vocabulary. That the development of story comprehension is connected with the growth of other language skills is perhaps not surprising, because – as pointed out in Section 1 – the child's understanding of stories on MAIN is assessed via the ability to *verbally express* this understanding.

Does this then mean that story comprehension scores are low in a child *because* his or her vocabulary scores (CLT) are low? Or that story comprehension scores are high *because* vocabulary scores are high? Whilst we cannot rule out this possibility, we believe that there need not be such a causal relation. Recall that the CLT vocabulary task measures knowledge of nouns and verbs, whilst the MAIN comprehension questions asks about goals and internal states of story characters. The questions and answers to these questions are not verbalised as nouns and verbs only. Low story comprehension scores and low vocabulary scores in one language may instead both be reflections of something else, namely low language proficiency in general. This, in turn, is likely to be influenced by a number of background factors, such as low exposure to the language, little high-quality interaction (e.g. few or no literacy-related or storytelling activities). As pointed out by Dickinson and Smith (1994, p. 118), story understanding may be fostered by the same cluster of experiences which foster vocabulary growth. In addition, "growth in vocabulary may itself lead to enhanced story comprehension skills" (Dickinson & Smith, 1994, p. 118).[25]

In the present study, no statistical tests were conducted to investigate the relationship between language input and story comprehension. What was done

25. As noted above, story comprehension moreover builds on the higher-order processes of inference generation and reasoning. With increasing cognitive maturity and expanding world knowledge, children are able to make more effective use of their limited attentional and working memory capacities during inference generation and reasoning, irrespective of their vocabulary skills or general proficiency in a particular language. The verbalisation of these inferences is, however, dependent on vocabulary and general language proficiency.

is that the information in the parental questionnaires of low-performing and high-performing children was systematically checked to see whether there were any patterns regarding input and whether there was any parental concern about language development (especially for the low-scoring children). Moreover, CLT scores were consulted, and it was seen that poor story comprehenders also performed poorly on the CLT vocabulary tasks in the respective language. Children who scored very low on narrative comprehension may not have had sufficient expressive vocabulary skills to formulate a comprehensible answer, and/or perhaps not even enough receptive vocabulary knowledge to fully understand the questions.

Questionnaire data on the children's socioeconomic and language backgrounds further indicated that poor story comprehenders in Turkish had only had limited exposure to their heritage language Turkish and were exposed to few, if any, literacy-related or storytelling activities in Turkish. Sometimes, limited exposure and limited Turkish home literacy went hand in hand with a third language being spoken in the home. Alternatively, the parents generally spoke Swedish with each other and with the child. The children's low performance in Turkish, coupled with information on language input and family background, suggests that these children might become 'receptive' or 'passive' bilinguals only, who are able to understand Turkish (to some extent) but not speak it in the future. For one child, there may also be concerns about a potential (as yet undiagnosed) developmental language disorder.

Children with exceptionally low Swedish MAIN comprehension scores also had very low Swedish vocabulary scores, as well as relatively little exposure to Swedish, either due to late onset of Swedish and/or limited preschool attendance. Even though the children with exceptionally low scores in Swedish grew up in a predominantly Turkish-speaking home environment, it is important to stress that Turkish at home is *not* a cause or predictor for low Swedish scores, far from it. The vast majority of our participants had predominantly Turkish-speaking home environments, but they did *not* perform particularly low in Swedish. Of course, these children also attended Swedish-medium (pre)schools.

Our case studies of individual children further showed that children with exceptionally good story comprehension not only had exceptionally high vocabulary results for their age, but also shared certain background characteristics: they were reported by their parents to be 'ahead of their peers', be 'a fast developer', or to 'know very many words', they had been extensively exposed to the language and frequently encountered language- and literacy-fostering activities in the home in the respective language (such as daily shared book reading, storytelling etc.). These children often grew up in mid-to-high-SES families. Whilst we have not investigated such factors statistically, the case studies nevertheless suggest that certain cultural habits of families may affect children's narrative comprehension (e.g. oral narratives, exposure to storybooks, cf. Sénéchal & LeFevre, 2002, Fiani et al., this volume).

Of course, home-based activities are not the only source of children's narrative experiences. Our participants spent a major part of the day in institutional daycare, and even though these preschools are bound by the same national Swedish curriculum to actively foster language, narrative, and social skills, there appears to be considerable variation in how the curriculum is put into practice. Variation in such preschool experiences may well affect children's narrative development (Dickinson & Smith, 1994; Boyd & Nauclér, 2001). Qualitative assessment of both preschool and home-based storytelling and literacy activities and their relation to story comprehension skill is something that warrants further investigation. Factors that boost or restrain children's story comprehension and the development of narrative skills would need to be studied further, so that we can better understand how to support language development in bilingual children.

Acknowledgements

We are grateful to Sibylle Dillström and Karin Koltay for their contributions regarding data collection, transcription, scoring development and data analysis. Thanks also to Ercan Aras, Mikaela Farkas-Behndig, Maria Johansson, Zübeyde Küçükgöl and Müjde Nordling for help with data collection. Many thanks to the children, parents, schools and preschools willing to participate; without them there would be no data and no results. We also thank two reviewers for helpful comments on an earlier version of this chapter.

Funding

Funded by: Swedish Research Council Grant.
Award ID: VR 421-2013-1309.
Award recipient: Ute Bohnacker.
Statement: This work was supported by a Swedish Research Council Grant (VR 421-2013-1309) to Ute Bohnacker.

References

Astington, J. W., & Pelletier, J. (2005). Theory of mind, language, and learning in the early years: Developmental origins of school readiness. In B. D. Homer & C. S. Tamis-LeMonda (Eds.), *The development of social cognition and communication* (pp. 312–352). New York, NY: Taylor & Francis, ProQuest Ebook Central.

Berman, R. A., & Slobin, D. I. (1994). Narrative structure. Chapter IIA. In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: A cross-linguistic developmental study* (pp. 39–84). New York, NY: Psychology Press.

Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. Hove: Psychology Press.

Bishop, D. V. M., & Adams, C. (1992). Comprehension problems in children with specific language impairment: Literal and inferential meaning. *Journal of Speech and Hearing Research*, 35(1), 119–129. https://doi.org/10.1044/jshr.3501.119

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

Bohnacker, U. (2013–2019). Language impairment or typical language development? Developing methods for linguistic assessment of bilingual children in Sweden. *Research project funded by the Swedish Research Council (VR 421-2013-1309), commonly referred to as BiLI-TAS* (Bilingualism, Language Impairment, Turkish, Arabic, Swedish).

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Bohnacker, U. (2018a). Guidelines for scoring CLT. Uppsala University, Version March 2018 (Unpublished material).

Bohnacker, U. (2018b). Guidelines for scoring macrostructure in MAIN. Uppsala University, Version November 2018 (Unpublished material).

Bohnacker, U. (2020). Bilingual development of Turkish-speaking children in Sweden. In L. Grenoble, P. Lane, & U. Røyneland (Eds.), *Linguistic Minorities in Europe Online*. Turkish. Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/lme.11697491

Bohnacker, U., Lindgren, J., & Öztekin, B. (2016). Turkish- and German-speaking bilingual 4-to-6-year-olds living in Sweden: Effects of age, SES and home language input on vocabulary production. *Journal of Home Language Research*, 1, 17–41. https://doi.org/10.16993/jhlr.26

Bohnacker, U., & Lindgren, J. (in press). MAIN story comprehension: What can we expect of a typically developing child? In S. Armon-Lotem & K. K. Grohmann (Eds.), *LITMUS in action: Cross-comparison studies across Europe*. Amsterdam: John Benjamins. Preprint retrieved from <http://uu.diva-portal.org/smash/get/diva2:1348899/FULLTEXT01.pdf> (18 June, 2020).

Boudreau, D. M. (2007). Narrative abilities in children with language impairments. In R. Paul (Ed.), *Language disorders from a developmental perspective: Essays in honour of Robin S. Chapman* (pp. 331–356). Mahwah, NJ: Lawrence Erlbaum Associates.

Boyd, S., & Nauclér, K. (2001). Sociocultural aspects of bilingual narrative development in Sweden. In L. Verhoeven & S. Strömqvist (Eds.), *Narrative development in a multilingual context* (pp. 129–151). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.23.05boy

Dickinson, D. K., & Smith, M. W. (1994). Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly*, 29(2), 105–122. https://doi.org/10.2307/747807

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: An investigation of their narratives and memory. *International Journal of Language and Communication Disorders*, 43(2), 201–218. https://doi.org/10.1080/13682820701366147

Ellis Weismer, S. (1985). Constructive comprehension abilities exhibited by language-disordered children. *Journal of Speech and Hearing Research*, 28, 175–184. https://doi.org/10.1044/jshr.2802.175

Dunn, L. M., & Dunn, D. M. (1997). *Peabody Picture Vocabulary Test – Third Edition*. Circle Pines, MN: American Guidance Service.

Fiani, R., Henry, G., & Prévost, P. (this volume). Narrative comprehension in Lebanese Arabic-French bilingual children. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension. Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins.
Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, 56. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.
Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–276). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-011
Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics*, 63. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.
Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing Cross-Linguistic Lexical Tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 196–240). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-010
Hayward, D., Schneider, P., & Gillam, R. B. (2009). Age and task-related effects on young children's understanding of a complex picture story. *The Alberta Journal of Educational Research*, 55(1), 54–72.
Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak–English children: A cross-language comparison. *Applied Psycholinguistics*, 37(1), 145–164. https://doi.org/10.1017/S0142716415000454
Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394. https://doi.org/10.1037/0033-295X.85.5.363
Klop, D., Visser, M., & Oosthuizen, H. (2012). Narrative profiles of 20 bilingual normally developing 6–7 year old South African children. Paper presented at 6th Meeting of COST Action IS0804. Berlin, Germany.
Koivistoinen, J. (2012). Att berätta på ryska och på svenska: Narrativ förmåga som instrument för språklig bedömning av flerspråkiga barn [Storytelling in Russian and Swedish: Narrative ability as a measure of language in multilingual children] (Unpublished M.Sc. thesis). Uppsala University.
Kunnari, S., & Välimaa, T. (this volume). Narrative comprehension in simultaneously bilingual Finnish-Swedish and monolingual Finnish children. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension. Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins.
Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4-6-year-old children. *Reading Research Quarterly*, 47(3), 259–282. https://doi.org/10.1002/rrq.020
Letts, C., & Leinonen, E. (2001). Comprehension of inferential meaning in language-impaired and language normal children. *International Journal of Language Communication Disorders*, 36(3), 307–328. https://doi.org/10.1080/13682820110045829

Lindgren, J. (2018). *Developing narrative competence: Swedish, Swedish–German and Swedish–Turkish children aged 4–6* (Studia Linguistica Upsaliensia 19). Uppsala: Acta Universitatis Upsaliensis.

Lindgren, J. (2019). Comprehension and production of narrative macrostructure in Swedish: A longitudinal study from age 4 to 7. *First Language*, 39(4), 412–432. https://doi.org/10.1177/0142723719844089

Lindgren, J., & Bohnacker, U. (this volume). Inferential comprehension, age and language: How Swedish-German bilingual preschoolers understand picture-based stories. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension. Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins.

Lynch, J. S., & van den Broek, P. (2007). Understanding the glue of narrative structure: Children's on- and off-line inferences about characters' goals. *Cognitive Development*, 22, 323–340. https://doi.org/10.1016/j.cogdev.2007.02.002

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology*, 29, 327–365. https://doi.org/10.1080/02702710802165416

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia*, 42, 1414–1434. https://doi.org/10.1016/j.neuropsychologia.2003.12.016

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, 37(1), 69–89. https://doi.org/10.1017/S0142716415000429

Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*. Published online 2 February 2018. https://doi.org/10.1080/13670050.2018.1434124

Öztekin, B. (2019). *Typical and atypical language development in Turkish-Swedish bilingual children aged 4–7* (Studia Linguistica Upsaliensia 25). Uppsala: Acta Universitatis Upsaliensis.

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, 38(1), 36–76. https://doi.org/10.1598/RRQ.38.1.3

Pavlenko, A. (2009). Narrative analyses. In L. Wei & M. Moyer (Eds.), *The Blackwell guide to research methods in bilingualism and multilingualism* (pp. 311–325). Oxford: Blackwell. https://doi.org/10.1002/9781444301120.ch18

Renfrew, C. E. (1969 [1997]). *Bus Story Test: A Test of Narrative Speech* (4th ed.). Bicester, Oxon: Winslow Press.

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37(1), 49–67. https://doi.org/10.1017/S0142716415000417

Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian–Russian children. *International Journal of Bilingualism*, 21(5), 617–635. https://doi.org/10.1177/1367006916643528

Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, 73(2), 445–460. https://doi.org/10.1111/1467-8624.00417

Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology*, 27(6), 960–974. https://doi.org/10.1037/0012-1649.27.6.960

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Norwood, NJ: Ablex.

Thompson, J. G., & Myers, N. A. (1985). Inferences and recall at ages four and seven. *Child Development*, 56, 1134–1144. https://doi.org/10.2307/1130228

Tompkins, V., Guo, Y., & Justice, L. M. (2013). Inference generation, story comprehension, and language skills in the preschool years. *Reading & Writing*, 26(3), 403–426. https://doi.org/10.1007/s11145-012-9374-7

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes* 15(1): 249–275. https://doi.org/10.1080/01638539209544812

Trabasso, T., Secco, T., & van den Broek, P. W. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Lawrence Erlbaum Associates.

Trabasso, T., Stein, N., Rodkin, P. C., Park Munger, M., & Baughn, C. R. (1992). Knowledge of goals and plans in the on-line narration of events. *Cognitive Development*, 7(2), 133–170. https://doi.org/10.1016/0885-2014(92)90009-G

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: disentangling bilingualism from language impairment* (pp. 301–330). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-013

Ukrainetz, T. A., Justice, L. M., Kaderavek, J. N., Eisenberg, S. L., Gillam, R. B., & Harm, H. M. (2005). The development of expressive elaboration in fictional narratives. *Journal of Speech, Language, and Hearing Research*, 48(6), 1363–1377. https://doi.org/10.1044/1092-4388(2005/095)

UNESCO Institute for Statistics (2012). *International Standard Classification of Education: ISCED 2011*. Montreal, Quebec: UIS.

Van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum Associates.

Westerveld, M. F. (2014). Emergent literacy performance across two languages: Assessing four-year-old bilingual children. *International Journal of Bilingual Education and Bilingualism*, 17(5), 526–543. https://doi.org/10.1080/13670050.2013.835302

Westerveld, M. F., Gillon, G. T., & Boyd, L. (2012). Evaluating the clinical utility of the Profile of Oral Narrative Ability for 4-year-old children. *International Journal of Speech-Language Pathology*, 14(2), 130–140. https://doi.org/10.3109/17549507.2011.632025

Widaman, K. F. (2006). Chapter 3: Missing data: What to do with or without them. In K. McCartney, M. R. Burchinal, & K. L. Bub (Eds.), *Best practices in quantitative methods for developmentalists. Monographs of the Society for Research in Child Development*, 71(3), Serial No. 285 (pp. 42–64).

Narrative comprehension in simultaneously bilingual Finnish-Swedish and monolingual Finnish children

Sari Kunnari and Taina Välimaa
University of Oulu, Finland

We analysed narrative comprehension in 5-to-6-year-old simultaneously bilingual Finnish-Swedish ($n = 16$) and monolingual Finnish children ($n = 16$) by using the Multilingual Assessment Instrument for Narratives (MAIN). We assessed mean total narrative comprehension scores for bilingual children in both of their languages and for monolingual children in Finnish, in both telling and retelling conditions. We compared bilingual and monolingual children's narrative comprehension in Finnish and analysed the association between comprehension and production. We also analysed the children's ability to answer different types of comprehension questions (i.e., questions probing goals, internal state terms, and questions requiring both the ability to draw inferences and to explain answers). We found no difference in total narrative comprehension scores for bilingual children between their two languages or between monolingual and bilingual children. This suggests language-independent narrative comprehension. We found no difference in narrative comprehension between telling and retelling and no correlation between narrative comprehension and production. However, we found a clear question type effect. Children performed better on questions probing goals or internal state terms, but questions that required both inferencing and ability to explain answers were very demanding. In conclusion, detailed analysis of narrative comprehension provides knowledge on how children create a coherent understanding of a story and utilise information in the comprehension process.

Keywords: bilingualism, narrative comprehension, preschool children, MAIN

1. Introduction

The basis of narration is established from an early age when children grow up with narratives presented in various situations (e.g. joint conversations, play situations, shared book readings and television watching, Lynch et al., 2008; Paris & Paris,

https://doi.org/10.1075/sibil.61.05kun

2003). Narrative comprehension involves a group of interrelated cognitive skills that allow a child to build a coherent representation of a story (Johnston, 2008; Lynch et al., 2008; Oakhill & Cain, 2007). Thus, the ability to comprehend narratives requires processes such as understanding and encoding the events, conceptually connecting different parts of narrative using prior knowledge and making inferences. All these skills are also essential to children's later social and academic achievements, especially for reading achievements (Lynch et al., 2008; Paris & Paris, 2003). Thus, in order to prevent the possible long-term effects of poor narrative comprehension, its assessment is crucial. It provides multifaceted information about the child's linguistics and cognitive abilities. There is a scarcity of knowledge regarding bilingual children's narrative comprehension skills, since a majority of the studies have focused on production rather than comprehension (Gagarina et al., 2012). In addition, research on bilingual children's narrative comprehension has provided inconsistent results. Some studies have shown clear differences in bilingual children's narrative comprehension between the languages they are acquiring (Gutiérrez-Clellen, 2002; Roch, Florit, & Levorato, 2016), whereas some studies indicate that acquiring several languages has no effect on the ability to comprehend narratives (Bohnacker, 2016; Kapalkova, Polišenská, Marková, & Fenton, 2016; Lindgren, 2018). Thus, the present study contributes to the previous literature by focusing on narrative comprehension in simultaneously bilingual Finnish-Swedish children and monolingual Finnish children between 5 and 6 years of age.

The base for narrative comprehension is the ability to connect mentally the different events of a story into a coherent whole. One crucial component in narrative comprehension is the role of causal connections among story events (e.g. physical and motivational causal connections; Kendeou et al., 2005; Lorch, Milich, & Sanchez, 1998; Trabasso & van den Broek, 1985; Trabasso, Secco, & van den Broek, 1984). That is, children have to determine the causes of a certain event and the effects of that event on the subsequent events in order to achieve a coherent understanding of a story. When children create these causal connections and make inferences between the events, they form mental network representations, which in turn direct children's comprehension of stories. When children are asked, for example, why-questions (e.g. why a certain event happened), they base their answers on these causal connections between events. Another important component in narrative comprehension is the organizational structure of the story, i.e. a hierarchically and causally organized structure of episodes. An episode consists of an entire behavioural sequence where units such us initiating events, goals, attempts, outcomes and a character's inner feelings are indicated (Stein & Glenn, 1979; Trabasso et al., 1984). The character's goal arises from story events and refers to states within the character, i.e. the desires and intentions of the character. A given goal may in turn motivate other action sequences and outcomes.

Narrative comprehension can be assessed through picture comprehension, listening comprehension or reading comprehension (Paris & Paris, 2003). According to a number of studies, similar processes seem to contribute to narrative comprehension across different domains (e.g. Kendeou et al., 2005; Lynch et al., 2008; van den Broek, 2005) suggesting that narrative comprehension is not domain specific. There might be differences, however, in what children comprehend through different domains. Through picture comprehension and listening comprehension it is possible to assess narrative comprehension skills independent of decoding skills as opposed to reading comprehension (Paris & Paris, 2003). The advantage of using pictorial stories in the assessment of the narrative comprehension of young children lies in the fact that children are often used to looking at pictures/pictorial stories, they are fun to look at, but still require a similar kind of cognitive skills as text-based stories. Assessment of narrative comprehension with pictorial stories minimizes the confounding factor of decoding skills and provides an opportunity to assess cognitive processes and linguistic abilities that are important for children's early reading development. Such assessment methods are highly needed to complement traditional assessment based on reading skills. In the present study, narrative comprehension of 5- to 6-year-old children was assessed with pictorial stimuli.

There has been a considerable amount of research on narrative comprehension in adults and older school-aged children, but research on preschool-aged children and early grades of elementary school-aged children is relatively rare. Previous research has shown that children's performance in narrative comprehension improves with age (e.g. Curenton, 2011; Hayward, Schneider & Gillam, 2009; Lynch et al., 2008; Paris & Paris, 2003), and there seem to be no gender differences in the ability to answer the comprehension questions (e.g. Lynch et al., 2008). In an itemized analysis of children's ability to answer different questions probing narrative comprehension, Curenton (2011) found that children performed better on questions addressing story characters' actions rather than their motives or intentions. Furthermore, van den Broek et al. (2005) noted that very young children are already able to identify connections between concrete and external events, whereas older children can increasingly identify connections among abstract and internal events (e.g. goals and feelings of characters). In addition, young children's inference making may be limited to identifying connections between individual events, whereas older children can connect groups of events such as episodes.

As regards the associations among early narrative comprehension skills, basic language and literacy skills and later reading comprehension skills, results are somewhat conflicting. In many studies, vocabulary has been related to narrative comprehension (Kendeou et al., 2005; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012; Potocki, Ecalle, & Magnan, 2013). However, van den Broek et al. (2005) found that early narrative comprehension skills are to a large extent independent from vocabulary. In addition, morphological and syntactic knowledge and sentence comprehension skills (Potocki et al. 2013) as well as sentence memory (Lepola et al. 2012) may predict narrative comprehension. Narrative comprehension may also be associated with some prereading skills such as phoneme segmentation (Paris & Paris, 2003). Some studies, however, have shown that comprehension of narratives develops independently from basic literacy skills such as phonemic awareness and letter and word identification (Kendeou et al., 2005; Lynch et al., 2008; van den Broek et al., 2005). Early narrative comprehension skills have also predicted later reading comprehension skills (e.g. Kendeou et al., 2005; van den Broek et al., 2005). Thus, assessment of narrative comprehension during preschool age gives important knowledge on children's competence to be shared, e.g. with parents and school personnel, and to be considered during school-start/schooling.

Research on narrative comprehension in bilingual children is scarce. Gutiérrez-Clellen (2002) has found that Spanish-English bilingual children (L1 Spanish and L2 English; age from 7 to 8 years; coming mainly from Mexican American background) exhibit greater comprehension of English than of Spanish stories and greater variability in the Spanish stories compared to the English stories. During the last few years, some studies using the recently developed Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015) have also addressed narrative comprehension of bilingual children. Studies on groups of mainly simultaneously Swedish-English, Swedish-German, or Swedish-Turkish and Turkish-German bilingual children have shown a clear age effect (Bohnacker, 2016; Lindgren, 2018; Maviş, Tuncer, & Gagarina, 2016), but no effect of language for narrative comprehension (Bohnacker, 2016; Lindgren, 2018). In addition, Maviş et al. (2016) found a task effect (i.e. children performed better in the "tell-after model" than in the "tell-no model" condition), but no gender effect. A study by Otwinowska, Mieszkowska, Białecka-Pikul, Opacki, and Haman (2018) on Polish-English bilingual children (exposed to English before the age of four; *M* = 12 months, range 0–48 months) showed a clear task effect (i.e. better comprehension in retelling than in telling conditions). They also showed a clear language effect with children performing better in English (i.e. the language of schooling and peer-to-peer interaction). The study by Roch et al. (2016) on sequentially bilingual Italian-English children showed a significant age and task effect (i.e. children scored higher in the story retelling task than in the telling task), but also a language effect (i.e. comprehension of L1 was better than L2). The study by Kapalková et al. (2016) on sequentially bilingual Slovak-English children, however, did not find any significant effect of language for narrative comprehension. Furthermore, deeper analysis regarding children's ability to comprehend specific macrostructural components (e.g. goals or internal states as initiating events) has shown that simultaneously bilingual children understand goals and internal states as initiating events well, regardless of language (Bohnacker, 2016). In contrast, a study on sequentially bilingual children showed that goals were the least well-comprehended components, while initiating events and reactions were clearly easier components for the children to comprehend (Kapalková et al., 2016).

Relatively little is known about young bilingual and monolingual children's comprehension of narratives. Studying narrative comprehension provides a unique opportunity to assess children's understanding of complex events. Analysis of narrative comprehension may also serve as a valuable contribution in the identification of developmental language delay or disorder and even early reading difficulties (cf. Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016). In the assessment of bilingual children, uniform and parallel assessment materials in both their languages are important in order to avoid biased results due to learning effect.

Thus, the purpose of the present study was to analyse narrative comprehension in Finnish-Swedish bilingual and Finnish monolingual children. The specific research questions were as follows: (1) Are there differences in total scores in narrative comprehension in Finnish-Swedish bilingual children in both their languages and in two elicitation tasks (telling and retelling)? (2) Are there differences in total scores in narrative comprehension between Finnish-Swedish bilingual and Finnish monolingual children? (3) Is there an association between narrative comprehension and production? (4) Are there differences in children's abilities to answer narrative comprehension questions probing different macrostructure components? In the present study, the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012, 2015; Kunnari & Välimaa, 2012) was applied, since it has been proven to be cross-culturally robust and it contains parallel elicitation tasks for use in several languages.

2. Method

2.1 Participants

A total of 32 children in an age range from 5;0 to 6;7 years [years; months] participated in this study: 16 simultaneously bilingual Finnish-Swedish children (bilingual group) and 16 monolingual Finnish children (monolingual group). The two groups were matched on age and gender. The parents completed a background questionnaire on demographic factors, on the patterns of language use in their family and the child's perceived comprehension and production skills in everyday situations for both languages (Gagarina et al., 2012). According to the parental questionnaires, the children were typically developing and had no history of speech or language disorders or use of speech therapy services. The children came from comparable socioeconomic backgrounds (i.e. about 90% of the parents of the children had higher education). The age range from 5;0 to 6;7 years was chosen, because according to the existing literature, narrative abilities are expected to develop clearly during this age (e.g. Muñoz, Gillam, Peña, & Gulley-Faehnle, 2003; Price, Roberts, & Jackson, 2006; Schneider, Hayward, & Dubé, 2006). From the sociocultural perspective, it is highly relevant to investigate children's narrative abilities at this age, that is, before they begin their formal education. Compulsory education (grade 1) starts at the age of 7 in Finland.

The bilingual group consisted of 16 simultaneously bilingual children (8 boys, 8 girls, mean age 5;8 years, $SD = 0.5$, min. 5;0, max. 6;6) recruited from bilingual Finnish-Swedish kindergartens. The bilingual children came from families where both languages had been spoken to them from birth. The input of Finnish and Swedish was approximately equal at the time of the study as assessed by the parents. The bilingual children were received a 46% ($SD = 17.7$) language input in Finnish and a 54% ($SD = 17.7$) input in Swedish. For 62% of the bilingual children, the overall daily exposure of both languages was equal and balanced. For the present study, Finnish and Swedish were chosen because they are the official national languages in Finland. Speakers of Swedish comprise only approximately 5.2% of the total population (Statistics Finland, 2018).

The monolingual group consisted of 16 monolingual Finnish-speaking children (8 boys, 8 girls, mean age 5;9 years, $SD = 0.6$, min. 5;1, max. 6;7) recruited from monolingual Finnish-speaking kindergartens. They came from homes where only Finnish was spoken. The monolingual group was matched with the bilinguals on the age of the child (independent samples t test, $t = -0.115$, $p = .91$) and gender (equal number of boys and girls in both groups). Monolingual Swedish-speaking children were not included in this study, because they would have been difficult to recruit due to the low ratio of Swedish-speaking population (Statistics Finland, 2018).

2.2 Procedure and design

Narrative comprehension and associations between narrative comprehension and production were assessed using the Finnish and the Swedish versions of MAIN (Bohnacker, 2012; Gagarina et al. 2012, 2015; Kunnari & Välimaa, 2012). MAIN includes narrative production and comprehension tasks. Narrative production is assessed with six-picture sequences matched in the overall story and episode structure in telling and retelling situations. Each story consists of a setting and three short episodes made up of a goal-attempt-outcome (GAO) sequence with comparable complexity, and internal state terms (ISTs) as initiating events or protagonists' reactions. Following the telling and retelling situations, the experimenter asks ten comprehension questions, while the whole story is visible to the child. The questions target the children's overall story comprehension and their inferencing ability concerning the thoughts, feelings, and intentions (goals) of the protagonists. Three questions probe the character's goals (e.g. *Why does the fox leap forward?*), three questions probe internal state terms as initiating events and reactions of the protagonists (e.g. *How does the fox feel?*), and four questions probe the ability to draw inferences and explain the answers (e.g. *Why do you think the fox is feeling bad/scared/hungry/disappointed etc.?*). Thus, the comprehension questions do not measure the children's ability to recall the story, but their ability to create causal connections and make inferences.

In the present study, each child was individually assessed in a quiet room at their kindergarten and the sessions were audio-recorded using an Olympus LS-11 recorder. At the beginning of the assessment session, the child was instructed to choose a story from one of three envelopes and not to let the examiner see the selected story. This controlled for effects of shared knowledge and joint attention. All of the stories were presented in a foldout fashion, initially showing the child the whole story with all six pictures. The examiner sat opposite to the child folding out the pictures so that only the child was able to see them. After this, the child was instructed to tell/retell the story, during which the pictures were unfolded in sets of two. The assessment procedure followed the same principles for telling and retelling, except for retelling condition, where the examiner first told the story to the child and then asked the child to retell it. After the child had told or retold the story, the child was asked comprehension questions. The examiner initiated narrative comprehension assessment by saying "Now I will ask you some questions about the story". The first question was always a warm-up question "Did you like the story?" and it was not scored. Thereafter, the examiner proceeded according to MAIN protocol.

For the bilingual group, narrative comprehension was assessed with two tasks (telling and retelling) in each language. The order of presentation was counterbalanced with regard to language (Finnish and Swedish) and the story (i.e., Baby Birds and Baby Goats for the telling task; Cat and Dog for the retelling task). For the monolingual group, narrative comprehension was assessed with two tasks (telling and retelling). For them, the order of presentation was counterbalanced with regard to the story. Both groups always began the assessment session with one of the telling tasks. The bilingual group was assessed by one (native/near native) bilingual examiner and the monolingual group by one native monolingual examiner. Both examiners spoke only in the language of the assessment during assessment session. The testing interval between the two assessment sessions of the bilingual children (Finnish/Swedish) was approximately 1 week.

The children's answers to comprehension questions were transcribed and scored using MAIN protocol (Bohnacker, 2012; Gagarina et al., 2012, 2015; Kunnari & Välimaa, 2012). The Swedish narratives were scored by the Swedish-speaking examiner and the Finnish narratives by the Finnish-speaking examiner. Children were given 1 point for each correct answer (maximum score = 10 per story). The decision on children's correct/incorrect responses was made based on more than 500 children's responses acquired during the developmental process of MAIN (Gagarina et al., 2012, 2015). For each story, examples of correct/incorrect responses were given in the questions and scoring sheet of narrative comprehension (see Gagarina et al., 2012, 2015). Ten per cent of the randomly selected comprehension question responses were scored by a second independent rater. Reliability was calculated by two different raters, with respect to children's answers to every question probing narrative comprehension. The point-to-point agreement for the rating was 96.7%. Narrative production, i.e. the story structure scores of the children, was drawn from previous narrative production analysis (cf. Kunnari et al., 2016). The production of story structure was analysed using the guidelines described in Gagarina et al. 2012 and 2015. The number of story structure elements was calculated: setting and an episode consisting of an internal state term (IST) as an initiating event, goal (G), attempt (A), outcome (O), and an IST as a reaction. In each story, children were awarded points for the setting statement and for each statement referring to any part of the episode. All of the stories included three episodes, so the children were awarded a maximum of 17 points for the production of story structure.

2.3 Statistical analysis

Mean total scores and standard deviations of narrative comprehension were first analysed for the bilingual group in Finnish and in Swedish and for the monolingual group in Finnish for telling and retelling conditions. Because narrative comprehension scores did not show normal distribution, nonparametric statistical tests were used systematically in the analysis.

To answer our first research question on narrative comprehension of the bilingual group in both their languages, total scores in narrative comprehension in Finnish and in Swedish were compared in each of the conditions by using the Wilcoxon Signed Ranks Test.

To answer our second research question on narrative comprehension of the bilingual and the monolingual groups in Finnish, the total narrative comprehension scores were compared using the Mann-Whitney U-test. The Wilcoxon Signed Ranks Test was conducted to analyse the differences between telling and retelling conditions.

To answer our third research question on associations between narrative comprehension and production, Spearman's rank order correlation coefficients were used to explore the associations between narrative comprehension and production in Finnish and in Swedish for the bilingual group and in Finnish for the monolingual group. Before statistical analysis, the raw scores for narrative comprehension (max. 10 points) and production (max. 17 points) were transformed into percentage scale (i.e., maximum score gave the percentage of 100%). This made the different scales comparable.

To answer our fourth research question on children's ability to answer questions probing comprehension of goals (maximum score = 3), internal state terms as initiating events and reactions of the protagonists (maximum score = 3), and to draw inferences and explain their answers (maximum score = 4), the total number of correct answers to these three types of narrative comprehension questions was analysed separately for both bilingual and monolingual groups. To analyse the ability of the bilingual children to answer the three types of narrative comprehension questions in both their languages, the Related-samples Friedman's two-way analysis of variance by ranks with the question as the within-subjects variable (i.e. goal; IST as initiating event or reaction; inferences and explanations) was conducted. Before statistical analysis, the raw scores were transformed into ratios (i.e. child's raw score divided by the maximum score for each question type) to make the scoring equal for the different question types. In order to control Type 1 error, adjustments for multiple comparisons were made using a Bonferroni method with $0.05/n$ as the threshold for significance. In the present study, n is the number of comparisons between the different question types. Similar analysis was conducted for the monolingual children in Finnish narratives. Differences between the bilingual and the monolingual groups in their ability to answer the three narrative comprehension question types in Finnish narratives were analysed using the Kruskal-Wallis Test with the question as a within-subjects variable (goal; IST as initiating event or reaction; inferences and explanations, INF&E) and the language group as a between-subjects variable (bilingual vs. monolingual).

3. Results

3.1 Narrative comprehension: Total score on comprehension questions

Table 1 illustrates mean comprehension scores for story telling and retelling tasks for the bilingual group in Finnish and in Swedish and for the monolingual group in Finnish. When narrative comprehension of the bilingual group was compared in both their languages, Wilcoxon Signed Ranks Test showed no differences between the total comprehension scores in Finnish and in Swedish either in telling ($Z = 34.5$, $p = .147$) or in retelling conditions ($Z = 26.5$, $p = .562$). Next, the two task types telling and retelling were compared for the bilingual group in both their languages. The results showed that there were no differences in the comprehension of Finnish ($Z = 50.5$, $p = .118$) or Swedish ($Z = 43.5$, $p = .572$) narratives between telling and retelling conditions.

When narrative comprehension in Finnish was compared between the bilingual and the monolingual groups, Mann-Whitney U-test showed no differences between the two groups in telling ($U = 159.0$, $p = .233$) or in retelling conditions ($U = 114.5$, $p = .582$) (Table 1). When narrative comprehension of the monolingual group was compared between telling and retelling conditions, Wilcoxon Singed Ranks Test showed no differences between the two conditions ($Z = 44.0$, $p = .590$).

Table 1. Means and standard deviations (SD, in brackets) for comprehension scores (max. 10) of the bilingual children in Finnish and in Swedish and for the monolingual children in Finnish

	Telling		Retelling	
	Finnish	**Swedish**	**Finnish**	**Swedish**
Bilingual group	6.3 (1.7)	5.2 (2.5)	5.8 (1.3)	5.6 (1.4)
Monolingual group	5.5 (2.1)		5.9 (1.6)	

3.2 Associations between narrative comprehension and production

To analyse associations between narrative comprehension and narrative production for the bilingual group in both their languages, and for the monolingual group in Finnish, Spearman rank order correlation coefficients were calculated. To enable the analysis of the associations between narrative comprehension and production, narrative production results were reproduced from previous analyses (see Kunnari et al., 2016). The results showed a mean score for the bilingual group of 4.6 (SD = 1.9) in the story telling task in Finnish and 5.4 (SD = 2.3) in the story telling task in Swedish. They scored on average 6.6 (SD = 1.8.) in the retelling task in Finnish and 6.4 (SD = 2.3) in the retelling task in Swedish. The monolingual group scored on average 6.9 (SD = 1.7) in the telling task in Finnish and 7.6 (SD = 2.1) in the retelling task in Finnish. As demonstrated in Table 2, no significant correlations existed for the bilingual group in Swedish or in Finnish narrative comprehension (total score) and production (story structure, i.e. number of macrostructure elements) in either story telling or retelling conditions. The same holds true for the monolingual group in Finnish narrative comprehension and production and in both task types (telling and retelling).

Table 2. Associations (Spearman's rank order correlation coefficients; *p*-values in brackets) between narrative comprehension and narrative production for the bilingual children in Finnish and Swedish, and for the monolingual children in Finnish in telling and in retelling conditions

				Swedish		**Finnish**	
				Telling	**Retelling**	**Telling**	**Retelling**
				Production	**Production**	**Production**	**Production**
Bilinguals	Swedish	Telling	Comp	.49 (.052)			
		Retelling	Comp		.09 (.721)		
	Finnish	Telling	Comp			.48 (.063)	
		Retelling	Comp				.35 (.191)
Monolinguals	Finnish	Telling	Comp			.18 (.517)	
		Retelling	Comp				.05 (.868)

Note. Comp = Comprehension score.

3.3 Children's performance on the different comprehension questions

To gain knowledge on narrative comprehension of the bilingual children, their ability to answer the different types of narrative comprehension questions in both their languages was analysed using the Related-Samples Friedman's two-way analysis of variance by ranks (Table 3 provides descriptive statistics). Results revealed a significant question type effect for the bilingual group in both their languages. Namely, bilingual children's performance on questions probing goals (Mean rank = 2.56) in Finnish narratives in telling conditions exceeded their performance on questions requiring both inferencing and the ability to explain the answers (Mean rank = 1.13) $F_R = 20.13$, $df = 2$, $p = < .001$. In addition, their performance on questions probing ISTs (Mean rank = 2.31) exceeded their performance on questions requiring both inferencing and the ability to explain the answers. This question type effect was also statistically significant after adjustment for multiple comparisons using a Bonferroni adjustment of 0.05/3 = 0.017 as the threshold for significance ($p = .018$). However, there was no significant difference in children's performance between questions probing ISTs and goals. In retelling condition, the results showed quite a similar pattern. Bilingual children's performance on questions probing goals (Mean rank = 2.81) and on questions probing ISTs (Mean rank = 2.13) exceeded the performance on questions requiring both inferencing and the ability to explain the answers (Mean rank = 1.06) $F_R = 25.67$, $df = 2$, $p < .001$. These differences were also statistically significant after adjustment for multiple comparisons ($p = .001$). However, there was no difference in children's performance between questions probing goals and ISTs.

Table 3. Means and standard deviations (SD, in brackets) for children's ability to answer questions probing comprehension of goals (maximum score = 3), internal state terms as initiating events and reactions of the protagonists (ISTs) (maximum score = 3) and to draw inferences and explain their answers (Inf&E) (maximum score = 4)

		Telling		Retelling	
		Finnish	**Swedish**	**Finnish**	**Swedish**
Bilinguals					
	Goal	2.5 (0.7)	2.2 (1.0)	2.6 (0.6)	2.2 (0.8)
	ISTs	2.1 (0.7)	1.7 (0.9)	1.8 (0.4)	1.9 (0.5)
	Inf&E	1.7 (0.8)	1.3 (1.1)	1.4 (0.8)	1.6 (0.7)
Monolinguals					
	Goal	2.0 (0.7)		2.9 (0.3)	
	ISTs	2.1 (1.0)		1.6 (0.7)	
	Inf&E	1.4 (1.0)		1.4 (0.8)	

In Swedish narrative comprehension, the question type effect was also clear. In the telling task, bilingual children performed better on questions probing goals (Mean rank = 2.56) and on questions probing ISTs (Mean rank = 2.31) than on questions requiring both inferencing and the ability to explain the answers (Mean rank = 1.13) $F_R = 21.19$, $df = 2$, $p < .001$. The question type effect was statistically significant even after adjustment for multiple comparisons ($p < .001$). No differences were found between questions probing goals and ISTs. Again, similar question type effect was evident in the Swedish retelling task. Children performed better on questions probing goals (Mean rank = 2.47) and on questions probing ISTs (Mean rank = 2.34) than on questions requiring both inferencing and the ability to explain the answers (Mean rank = 1.19) $F_R = 17.32$, $df = 2$, $p < .001$). There were no differences between questions probing goals and ISTs.

Monolingual children's ability to answer the different narrative comprehension questions was analysed in Finnish narratives in telling and retelling conditions. The results showed a question type effect, but the patterns were somewhat different from bilingual children. In telling condition, monolingual children performed better on questions probing goals (Mean rank = 2.28) and ISTs as initiating events or reactions of the protagonists (Mean rank = 2.53) than on questions requiring both inferencing and the ability to explain the answers (Mean rank = 1.19) $F_R = 18.35$, $df = 2$, $p < .001$. After adjustment for multiple comparisons, the difference was also statistically significant. Quite in line with the bilingual children, there were no statistically significant differences in monolingual children's abilities to answer questions probing goals or ISTs. However, the retelling task showed somewhat different results. After adjustment for multiple correlations, there were no differences in children's performance on questions probing goals (Mean rank = 2.97) and ISTs (Mean rank = 2.00), or between questions probing ISTs and questions requiring both inferencing and the ability to explain the answers. The only difference remained between goals and those that required the ability to explain the answers. Monolingual children's performance on questions probing goals (Mean rank = 2.97) exceeded the performance on questions requiring both inferencing and the ability to explain the answers (Mean rank = 1.03), $F_R = 31.0$, $df = 2$, $p < .001$. This difference was significant also after adjustment for multiple comparisons.

The possible differences in the ability of the bilingual group and the monolingual group to answer the three types of narrative comprehension questions in Finnish narratives were also analysed in order to consider the effect of language group on children's performance. Results revealed that language group (bilingual vs. monolingual) had mainly no effect for children's performance on questions probing ISTs as initiating events or reactions of the protagonists (X^2 (1) = .17, $p = .685$) or questions requiring both inferencing and explaining the answers (X^2 (1) = .77, $p = .379$) in telling conditions. The only difference was their performance on questions probing goals: the bilingual children performed better than the monolingual children did (X^2 (1) = 4.28, $p = .039$). In retelling condition, there were no differences between bilingual and monolingual children in their performance on questions probing goals (X^2 (1) = 3.26, $p = .071$), ISTs as initiating events or reactions of the protagonists (X^2 (1) = .29, $p = .588$) or inferencing and explaining the answers (X^2 (1) = .08, $p = .781$). This indicates that children's performance differences in the three types of comprehension questions were not due to bilingual condition.

4. Discussion

This study examined narrative comprehension of simultaneously bilingual Finnish-Swedish and monolingual Finnish 5- to 6-year-old children. Children's narrative comprehension was compared for language (Finnish vs. Swedish), group (monolinguals vs. bilinguals), and task (telling vs. retelling) and associations between narrative comprehension and production were analysed. In addition, the study analysed children's ability to answer the three types of narrative comprehension questions (i.e. questions probing goals, internal state terms as initiating events and reactions, and questions requiring the ability to draw inferences and explain the answers). The study showed no differences in narrative comprehension of simultaneously bilingual Finnish-Swedish children between their two languages for either telling or retelling conditions. This study showed no differences between monolingual and bilingual children in Finnish narrative comprehension and no differences between the two tasks. Narrative comprehension seemed not to be associated with narrative production. Instead, this study showed significant differences in the children's ability to answer different narrative comprehension questions. Questions probing comprehension of goals and ISTs were easiest for both groups of children. Conversely, questions that required both inferencing and the ability to explain the answers were the most difficult for both groups of children.

4.1 Language comparison

The finding that there were no statistically significant differences in narrative comprehension in simultaneously bilingual Finnish-Swedish children between their two languages for either telling or retelling conditions suggests a fairly balanced competence in both languages. As such, our results imply that at the age of 5 to 6 years, these bilingual children had already established and were using an underlying mental schema for narrative comprehension, and were able to convey this in their answers quite in line with monolingual children (cf., Shapiro & Hudson, 1991; Stein & Albro, 1997; Trabasso & Rodkin, 1994; Trabasso & Stein, 1994). Our results accord with those by Bohnacker (2016) and Lindgren (2018), who found no language effect for narrative comprehension in mainly simultaneously bilingual Swedish-English, Swedish-German and Swedish-Turkish children. In addition, Kapalková et al. (2016) reported no language effect for narrative comprehension in sequentially bilingual Slovak-English children. The results of the current study are somewhat in contrast with the findings of Gutiérrez-Clellen (2002) and Roch et. al. (2016) who reported a clear language effect. Gutiérrez-Clellen (2002) showed that sequentially bilingual children exhibited better narrative comprehension in L2 and Roch et al. (2016) reported better comprehension in L1. These differences in results suggest different balance between the two languages of the bilingual children. Indeed, in Roch et al. (2016), the children were sequentially bilingual Italian-English children. Since the same methodology was used in Roch et al. and in the present study, the contrasting results most likely arise from the different balance between the two languages (i.e. simultaneous vs. sequential bilingualism). The results of this study confirm the recent findings that narrative macrostructure is a relatively unbiased measure to assess language abilities of bilingual children (e.g. Boerma et al., 2016; Hipfner-Boucher et al., 2015). However, since some studies have shown a language effect for narrative macrostructure (e.g. Gutiérrez-Clellen, 2002; Roch et al., 2016), it is reasonable to pose a question about the level of language proficiency needed for conveying the elements of narrative macrostructure with the languages the child is using.

4.2 Group comparison

In our study, no group effect (bilingual vs. monolingual) was found for narrative comprehension among 5- to 6-year-old bilingual and monolingual children. In many of the previous studies bilingual children's narrative comprehension has been analysed in both of the languages the children were acquiring (Bohnacker, 2016; Gutiérrez-Clellen, 2002; Lindgren, 2018; Roch et. al., 2016). Only the study by Lindgren (2018) provides a comparison between monolingual and bilingual children. In her study, no differences were found between Swedish monolingual and Swedish-German and Swedish-Turkish bilingual children in three out of four narrative tasks. In one of the tasks (Baby Goats), a significant difference was evident between the three language groups. The Swedish-Turkish bilinguals performed lower than the two other language groups. As such, our findings are somewhat in contrast with the one by Lindgren (2018).

One possible explanation for the lack of a difference in narrative comprehension between the bilingual and the monolingual children of the present study may stem from the fact that the simultaneously bilingual children could transfer their knowledge on narrative macrostructure between both their languages (Finnish and Swedish), because the macrostructure is similar across the languages (cf., Squires et al., 2014). In narratives, the bilinguals may share a conceptual base across both languages. Bearing in mind that the bilingual children of this study were all simultaneously bilingual, one could hypothesise that such transfer occurs relatively early in development diminishing the possible differences between bilingual and monolingual children. Another possible explanation may lie in the cultural similarities between Finnish-Swedish bilinguals and Finnish monolinguals. They reside in relatively similar cultural backgrounds, day care and preschool systems, where the story telling traditions are quite parallel providing the children with comparable model of narration and enabling them to comprehend the narrative macrostructure in a comparable manner in both of the languages (cf., Mäkinen, Gabbatore, Loukusa, Kunnari, & Schneider, 2019). Thus, the bilingual children's competence to utilise the mental schema of narratives in Finnish seems to be as proficient as the ability of the monolingual children (cf. Shapiro & Hudson, 1991; Stein & Albro, 1997; Trabasso & Rodkin, 1994; Trabasso & Stein, 1994). Detailed knowledge on the possible differences between bilingual and monolingual children's narrative comprehension abilities in the language used in education (e.g. English vs. Spanish; Finnish vs. Swedish; Swedish vs. Turkish) may provide essential implications for children's later social and academic achievements, especially for reading achievements (Lynch et al., 2008; Paris & Paris, 2003). However, if such a difference is discovered early enough, support may be given to ease children's school achievement and literacy.

4.3 Associations between narrative comprehension and production

No significant correlations existed between comprehension and production for the bilingual and monolingual children in either story telling or retelling conditions. This finding is in contrast with Roch et al. (2016) who found some correlation between narrative comprehension and production of Italian-English sequentially bilingual children, especially in English narratives. In addition, some earlier studies on monolingual and bilingual children have shown at least some association (e.g. Cain, 2003; Roch et al., 2016). However, in our study, for bilingual children in the telling condition in Swedish and in Finnish (see Table 2), the values of correlation coefficient were close to the critical values of moderate and statistically significant association for a sample size of 16 participants (r-values close to 0.5; p -values close to 0.05; e.g. Nummenmaa, 2009). The analysis nevertheless failed to reach statistical significance. One can ask whether the statistical analysis would have shown slightly different results with a larger sample size. In addition, it is important to note that the values of correlation coefficient were close to critical values of statistical significance in telling condition, but not in retelling condition. In a retelling task, the ability to answer narrative comprehension questions may pose demands also on children's memory, because they have heard the model story before answering the comprehension questions. Indeed, association between children's working memory and narrative comprehension has been shown in some previous studies (e.g. Dodwell & Bavin, 2008). Thus, more research is warranted in order to analyse the associations between narrative comprehension and production with slightly larger sample sizes, and in both the task types telling and retelling.

4.4 Task comparison

In our study, narrative comprehension was analysed separately in telling and retelling conditions. Such task comparison was enabled by the developmental work of MAIN (Gagarina et al., 2012, 2015), which includes parallel stories to be used in these two task types and in both of the languages the child is acquiring. Our results showed that there were no differences in the comprehension of narratives in telling or retelling conditions, for bilingual or monolingual children. There is scarcity in the current literature on the comparison of narrative comprehension between the two task types (Maviş et al., 2016; Roch et al., 2016). Our results are in concord with those of Maviş et al., (2016) who also found no difference in narrative comprehension between these two task types in simultaneously bilingual Turkish-German children. In contrast, Roch et al. (2016) showed that the sequentially bilingual children of their study performed better in the story retelling than in the telling task. This difference was evident in narrative comprehension as well as in production. One possible explanation for the differences may lie in the bilingual status of the children. In the present study and the one by Maviş et al., (2016), children were simultaneously bilingual. Thus, the results indicate that bilingual children's narrative comprehension in both languages may develop equally well, especially when children are acquiring both languages simultaneously. Indeed, in the study by Roch et al. (2016), children were sequentially bilingual and they scored lower especially in L2 (English) narrative comprehension in the telling condition. In narrative comprehension in telling condition, children do not receive a model story narrated by the experimenter as in the retelling condition, before the narrative comprehension questions are asked. It may well be that story telling and retelling tasks require different processing load, and in a retelling task children may take advantage of the previously heard model story. Thus, the contradictory results of possible differences in narrative comprehension between the two task types (telling and retelling) highlight the need to assess narrative comprehension in both conditions and the role of working memory in narrative comprehension.

4.5 Comprehension of different macrostructural components

To gain knowledge of children's ability to answer questions probing the different macrostructural components, children's answers to the questions were analysed in more detail. Results revealed a significant question effect for the bilingual group in Finnish and Swedish, and for the monolingual children in Finnish. Questions probing comprehension of goals (e.g. *Why does the fox leap forward?*) proved to be the easiest comprehension questions for both groups of children. Moreover, questions probing ISTs were also relatively easy for the children, even though they proved to be somewhat more difficult than questions probing goals. Our results accord with the ones by Bohnacker (2016) who also found that simultaneously bilingual children understood goals and internal states as initiating events well, regardless of language. Our results also partly accord with those by Curenton (2011) who found that the monolingual children in her study performed better on questions addressing the character's actions versus his motives/intentions.

One very interesting finding of the present study was that questions in which children were required both to draw inferences and to explain their own answers (e.g. *Why do you think the fox is feeling bad?*) were the most demanding comprehension questions for both groups of children. The 5–6-year old children of this study could be able to answer the questions probing ISTs as reaction correctly (e.g. *How does the fox feel?*) but could fail to explain his/her answer (e.g. *Why do you think the fox is feeling bad?*). These questions require both the ability to understand causal relations between events and the ability to explain them. It has previously been shown that consciousness questions (i.e. questions addressing a character's beliefs) may be demanding (e.g. Curenton, 2011). It has also been shown that giving explanations about inferences made during interpretation is a very demanding task for children (Letts & Leinonen, 2001). Explaining one's own answers requires many cognitive skills, such as an ability to e.g., distinguish between cause and result, and between pieces of evidence and conclusion. Thus, explanations can reveal children's awareness of the information that they have utilised in the comprehension process. According to recent findings, the ability to explain one's answers develops clearly during preschool and elementary school. For example, 5–6-year-old children have been able to explain approximately 70% of their answers correctly, and 8-year-old children already 80% (Loukusa, Mäkinen, Gabbatore, Laukkanen-Nevala, & Leinonen, 2017). In future, detailed analysis of children's incorrect answers to comprehension questions could provide additional data about the development of narrative comprehension and shed light on the children's awareness of the information used in comprehension.

5. Conclusions

The present study highlights that MAIN (Bohnacker, 2016; Gagarina et al., 2012, 2015; Kunnari & Välimaa, 2012) is a functional, easy and convenient tool in the assessment of narrative comprehension both in monolingual and bilingual children. It provides multifaceted information on narrative comprehension. Thus, it may serve as a complementary assessment tool in the identification of language competence, language delay or even language disorder (cf., Boerma et al., 2016). The present study provides some evidence that narrative comprehension is largely language-independent in cases where children are acquiring both languages simultaneously from birth and show equal narrative comprehension competence in both of their languages. This may also indicate fairly balanced bilingualism. However, further studies with larger sample sizes are warranted to confirm this finding. Because comprehension skills are essential to children's later social and academic achievements, especially for reading achievements (Johnston, 2008; Lynch et al., 2008; Oakhill & Cain, 2007), information on narrative comprehension abilities of bilingual children in both their languages during preschool years is essential even in the case of simultaneous bilingualism. Only through assessment, can we be sure of the level of narrative comprehension competence and offer support if needed.

In addition, the present study implies that questions probing different elements of episodes (e.g. goals) and questions that require inferencing, or require both the ability to draw inferences and to explain one's answers, pose different requirements for the narrative comprehension competence and inferencing skills. Furthermore, questions that require the ability to explain one's own answers can also reveal children's awareness of the information that they have utilised in the comprehension process. Thus, detailed analysis of children's narrative comprehension competence gives in depth knowledge of how children create causal connections, make inferences between events, create a coherent understanding of a story, and utilise information in the comprehension process.

Acknowledgements

We thank Riikka Lankila and Milla Bucht for their assistance with data collection and the children and their families for participating in this study.

Funding

Funded by: COST Action IS0804.
Funded by: Academy of Finland.
Statement: The research reported in this article was supported by the COST Action IS0804 and by the Academy of Finland.

References

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

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

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology*, 21, 335–351. https://doi.org/10.1348/026151003322277739

Curenton, S. M. (2011). Understanding the landscapes of stories: The association between preschoolers' narrative comprehension and production skills and cognitive abilities. *Early Child Development and Care*, 181(6), 791–808. https://doi.org/10.1080/03004430.2010.490946

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: An investigation of their narratives and memory. *International Journal of Communication and Language Disorders*, 43(2), 201–218. https://doi.org/10.1080/13682820701366147

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives (MAIN). *ZAS Papers in Linguistics, 56*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–276). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-011

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, 13(2), 175–197. https://doi.org/10.1016/S0898-5898(01)00061-4

Hayward, D. V., Schneider, P., & Gillam, R. B. (2009). Age and task-related effects on young children's understanding of a complex picture story. *The Alberta Journal of Educational Research*, 55(1), 54–72.

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenlee, J., Pelletier, J., & Girolametto, L. (2015). Narrative abilities in subgroubs of English language learners and monolingual peers. *International Journal of Bilingualism*, 19(6), 677–692. https://doi.org/10.1177/1367006914534330

Johnston, J. (2008). Narratives: Twenty-five years later. *Topics in Language Disorders*, 28(2), 93–98. https://doi.org/10.1097/01.TLD.0000318931.08807.01

Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak-English children: A cross-language comparison. *Applied Psycholinguistics*, 37(1), 145–164. https://doi.org/10.1017/S0142716415000454

Kendeou, P., Lynch, J. S., van den Broek, P., Espin, C. A., White, M. J., & Kremer, K. E. (2005). Developing successful readers: Building early comprehension skills through television viewing and listening. *Early Childhood Education Journal*, 33(2), 91–98. https://doi.org/10.1007/s10643-005-0030-6

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish-Swedish children. *Applied Psycholinguistics*, 37(1), 123–144. https://doi.org/10.1017/S0142716415000442

Kunnari, S., & Välimaa, T. (2012). Part II. MAIN: Finnish version (suomi). Multilingual Assessment Instrument for Narratives (MAIN). *ZAS Papers in Linguistics, 56*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4–6-year-old children. *Reading Research Quarterly*, 47(3), 259–282. https://doi.org/10.1002/rrq.020

Letts, C., & Leinonen, E. (2001). Comprehension of inferential meaning in language-impaired and language normal children. *International Journal of Language and Communication Disorders*, 36(3), 307–328. https://doi.org/10.1080/13682820110045829

Lindgren, J. (2018). *Developing narrative competence. Swedish, Swedish-German and Swedish-Turkish children aged 4–6* (Studia Linguistica Upsaliensia 19). Uppsala: Acta Universitatis Upsaliensis.

Lorch, E. P., Milich, R., & Sanchez, P. (1998). Story comprehension in children with ADHD. *Clinical Child and Family Psychology Review*, 1(3), 163–178. https://doi.org/10.1023/A:1022602814828

Loukusa, S., Mäkinen, L., Gabbatore, I., Laukkanen-Nevala, P., & Leinonen, E. (2017). Understanding contextual and social meaning in typically developing Finnish-Speaking four-to-eight-year-old children. *Psychology of Language and Communication*, 21(1), 408–428. https://doi.org/10.1515/plc-2017-0020

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other reading skills. *Reading Psychology*, 29(4), 327–365. https://doi.org/10.1080/02702710802165416

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, 37(1), 69–90. https://doi.org/10.1017/S0142716415000429

Muñoz, M. L., Gillam, R. B., Peña, E. D., & Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. *Language, Speech, and Hearing Services in Schools*, 34(4), 332–342. https://doi.org/10.1044/0161-1461(2003/027)

Mäkinen, L., Gabbatore, I., Loukusa, S., Kunnari, S., & Schneider, P. (2019). A comparison of picture-based narratives by Finnish, Italian and English-speaking children. *Early Education and Development*. https://doi.org/10.1080/10409289.2019.1666446

Nummenmaa, L. (2009). *Käyttäytymistieteiden tilastolliset menetelmät*. Helsinki: Tammi.

Oakhill, J. V., & Cain, K. (2007). Introduction to comprehension development. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 3–40). New York, NY: Guilford.

Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, Online First, 1–25. https://doi.org/10.1080/13670050.2018.1434124

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, 38(1), 36–76. https://doi.org/10.1598/RRQ.38.1.3

Potocki, A., Ecalle, J., & Magnan, A. (2013). Narrative comprehension skills in 5-year-old children: Correlational analysis and comprehender profiles. *The Journal of Educational Research*, 106(1), 14–26. https://doi.org/10.1080/00220671.2012.667013

Price, J. R., Roberts, J. E., & Jackson, S. C. (2006). Structural development of the fictional narratives of African American preschoolers. *Language, Speech, and Hearing Services in Schools*, 37(3), 178–190. https://doi.org/10.1044/0161-1461(2006/020)

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37(1), 49–67. https://doi.org/10.1017/S0142716415000417

Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology*, 27(6), 960–974. https://doi.org/10.1037/0012-1649.27.6.960

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech–Language Pathology and Audiology*, 30(4), 224–238.

Stein, N. L., & Albro, E. R. (1997). Building complexity and coherence: Children's use of goal structured knowledge in telling stories. In M. Bamberg (Ed.), *Narrative development: Six approaches* (pp. 5–44). Mahwah, NJ: Lawrence Erlbaum Association.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives* (pp. 53–120). Norwood, NJ: Ablex.

Statistics Finland. (2018). *Statistics Finland*. Retrieved from <https://www.stat.fi/index_en.html> (26 June, 2020).

Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language & Communication Disorders*, 49(1), 60–74. https://doi.org/10.1111/1460-6984.12044

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24(5), 612–630. https://doi.org/10.1016/0749-596X(85)90049-X

Trabasso, T., & Rodkin, P. C. (1994). Knowledge of goal/plans: A conceptual basis for narrating "Frog where are you?" In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: A cross-linguistic developmental study* (pp. 85–106). New York, NY: Psychology Press.

Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Mahwah, NJ: Lawrence Erlbaum Associates.

Trabasso, T., & Stein, N. L. (1994). Using goal–plan knowledge to merge the past with the present and the future in narrating events online. In M. M. Haith, J. B. Benson, R. J. Roberts , Jr., & B. F. Pennington (Eds.), *The development of future-oriented processes* (pp. 323–349). Chicago, IL: University of Chicago Press.

Van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M. J., & Lorch, M. P. (2005). Assessment of comprehension abilities in young children. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum Associates.

Narrative comprehension by Croatian-Italian bilingual children 5–7 years old

The role of receptive vocabulary and sentence comprehension

Maja Roch and Gordana Hržica
University of Padua / University of Zagreb

Objective. This study compares L1 and L2 receptive vocabulary, receptive grammar and narrative comprehension skills in Croatian-Italian bilingual children. Moreover, the study aims to find out to what extent receptive vocabulary and receptive grammar (sentence comprehension) predict narrative comprehension skills.

Method. Thirty Croatian-Italian bilinguals 5–7 years old were assessed in L1 (Croatian) and L2 (Italian) narrative comprehension (MAIN), receptive vocabulary (PPVT) and sentence comprehension (TROG).

Results. Children performed better in their L1 than in their L2 on all three measures of comprehension. Narrative comprehension correlated with the two linguistic skills in L1 but weakly in L2. Each measure correlated only with itself between L1 and L2. Regression analyses showed that sentence comprehension contributed substantially to narrative comprehension in L1 and L2, while receptive vocabulary contributed substantially only in L1.

Conclusions. Narrative comprehension is differently predicted by language skills in L1 and in L2. The contribution of language skills is monolingual-like in L1 narrative comprehension (cf. Florit, Roch, & Levorato, 2011); in L2 narrative comprehension vocabulary provides a weak contribution. The results are discussed both theoretically, in terms of the possible mechanism underlying narrative comprehension in bilingual speakers and practically, in terms of bilingual language development.

Keywords: bilingual children, receptive vocabulary, receptive grammar, narrative comprehension

https://doi.org/10.1075/sibil.61.06roc

1. Introduction

Societies have become increasingly multilingual in recent decades, and the number of children speaking more than one language has grown exponentially (Paradis, Geneese, & Crago, 2011). This has helped motivate research into the trajectories of language development in preschool bilingual children, which has established that bilingual speakers may show linguistic delays compared to their monolingual peers, particularly bilingual speakers not exposed to both languages from birth (Bonifacci, Barbieri, Tommasini, & Roch, 2018; Hoff & Core, 2015). This reflects that learning two languages takes longer than learning one and raises the question of what could be done during early linguistic development to ensure better linguistic outcomes prior to formal school instruction.

Vocabulary and syntactic development in bilingual children has received attention in the literature, while narrative comprehension has been less studied. The present work addresses the contribution of vocabulary and sentence comprehension to narrative comprehension in bilingual Croatian (L1)-Italian (L2) children.

Narratives are pervasive in children's lives from their earliest language experiences. Young children experience narratives through shared book reading at home or in educational settings. Thus, narrative competence emerges early during development, well before a child is able to read, yet most research into narrative comprehension has focused on school-aged children (Cain & Oakhill, 2007) and has neglected developmental trajectories of narrative competence during preschool years (Florit, Levorato, Roch, & Altoè, 2009; Florit, Roch, & Levorato, 2011, 2013, 2014). Research is urgently needed into the mechanisms of preschool development of narrative comprehension, in monolinguals and bilinguals, since it may be the best predictor of later reading comprehension (Kendeou, van den Broek, White, & Lynch, 2007; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012).

1.1 Assessing narrative skills

Narrative competence includes skills in both narrative production and narrative comprehension. Assessment of both skill sets can provide insights into a wide range of linguistic abilities as well as cognitive and pragmatic skills. Asking children to produce a brief sample narration provides rich data on story structure, structural complexity, internal state language, cohesion, morpho-syntax, and lexical diversity and productivity (Bohnacker, 2016).

Two distinct but interrelated areas underlie competence in narrative discourse: macrostructure and microstructure (Justice et al., 2006). Macrostructure concerns the higher-order mental organization of narratives, where content is connected through a global organization structure following an underlying narrative schema. Microstructure, which is much more language-specific, concerns the linguistic forms used in the narrative discourse (Gagarina, Klop, Tsimpli, & Walters, 2016; Iluz-Cohen & Walters, 2012; Simon-Cereijido & Gutiérrez-Clellen, 2009). Most studies of children's' narrative skills have analyzed production but not comprehension.

Narrative comprehension is usually measured by asking children to answer questions about a narrative that was read aloud to them or shown to them as a series of pictures. The questions target information explicitly narrated in the story as well as information that must be inferred, such as characters' motives and mental states. Narrative comprehension tasks therefore address both linguistic and cognitive abilities.

The understanding of picture stories is verbally mediated, and the comprehension of a story read aloud involves similar processes as comprehension of a story presented in pictures (Bishop & Adams, 1992). During narrative comprehension, a child attempts to construct a meaning-based representation of the narrative through a number of processes at the word, sentence, and text levels (Oakhill & Cain, 2007; Perfetti, Landi, & Oakhill, 2005), making it a very complex task. At the word and sentence levels, children build linguistic information about the narrative and store it in verbal working memory. As the child follows the story, he/she transforms the pictures into a verbal representation of their contents (Bishop & Adams, 1992). Words and their meanings are integrated into the complex meaning of sentences. Individual meanings of sentences are then put together into a more global representation of the story content (Cain, Oakhill, & Bryant, 2004; Cain & Oakhill, 2009). At the same time, the child deploys so-called "higher-level" cognitive processes such as inference-making in order to identify characters and their motives, follow the story structure and understand explicit and implicit information in the story, by integrating information contained in the story with previous world knowledge (e.g. Florit et al., 2014; Kintsch, 1994; Oakhill & Cain, 2007; Silva & Cain, 2015; van den Broek, Kendeou, Kremer, Lynch, Butler, White, & Purgzles Lorch, 2005).

1.2 Narrative comprehension in monolingual children

The relative contribution of the various components to narrative comprehension has been investigated in order to identify predictors of monolingual narrative comprehension. Most studies in this area have examined the comprehension of narratives read aloud (e.g., Florit et al., 2011; Kendeou et al., 2009; Kendeou, Bohn-Gettler, White, & van den Broek, 2008; Lepola et al., 2012; Oakhill & Cain, 2007; Roth, Speece, & Cooper, 2002). Only a few studies have examined the comprehension of narratives in pictures (Bishop & Adams, 1992; Bohnacker, 2016; Roch et al., 2016). Regardless of the narrative type, linguistic knowledge forms the essential precondition for complete understanding: adequate linguistic processing allows the child to process non-linguistic aspects of the narrative.

This linguistic knowledge can be measured mainly in terms of receptive vocabulary and sentence comprehension. Receptive vocabulary is one of the best predictors of oral narrative comprehension by preschool children (Kim, 2016). For example, receptive vocabulary as measured by the (PPVT) positively correlated with narrative competence for the pictured story "Frog, where are you?" (Korecky-Kröll, Dobek, Blaschitz, Sommer-Lolei, Boniecki, Uzunkaya-Sharma, & Dressler, 2019). Children acquire new words primarily through exposure to narratives (Cain, Oakhill, Barnes, & Bryant, 2001; Roth et al., 2002), and their comprehension of narratives increases as they become more familiar with the words contained therein (Seigneuric & Ehrlich, 2005).

Sentence comprehension in relation to narrative comprehension has been less frequently studied than receptive vocabulary. The relationship between narrative and sentence comprehension has been shown to be much weaker than between oral narrative comprehension and vocabulary (Long, Oppy, & Seely, 1997; Oakhill & Yuill, 1996). On the other hand, some studies have suggested that difficulties in narrative comprehension are related to poor processing of sentences (Nation, Clarke, Marshall, & Durand, 2004; Stothard & Hulme, 1992). This literature seems to suggest that during comprehension of oral narratives, the child must analyze the syntactic structure of sentences and identify their meaning in order to put them together into a mental representation of the entire story. If the child struggles with sentence comprehension, fewer resources are available for cognitive processes that integrate information about story content into a mental representation.

Bishop and Adams (1992) found a high correlation between sentence comprehension as measured with the (TROG) and comprehension of oral and pictured narratives. Those authors interpreted their results to indicate that understanding of a text requires constructive processing, rather than passive reception. The child builds a mental representation from a sequence of propositions, even when such propositions are presented nonverbally.

1.3 Narrative comprehension in bilingual children

Bilingual children are regularly exposed to more than one language in everyday life (De Lamo White & Jin, 2011), so they generally receive less exposure to each language than their monolingual peers (Hoff et al., 2012). This impacts on their language abilities, including production and comprehension of narratives. In recent years, narratives have been used for assessing bilingual language development during preschool years and for establishing a relationship between bilingual exposure and language delay (Cleave, Girolametto, Chen, & Johnson, 2010; Iluz-Cohen & Walters, 2012; Pesco & Kay-Raining Bird, 2016; Simon-Cereijido & Gutiérrez-Clellen, 2009). The literature on pre-schoolers indicates that bilingual and monolingual children show similar narrative competence as far as macrostructure is concerned, while bilinguals tend to struggle with microstructure (Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Bohnacker, 2016; Bonifacci et al., 2018; Rodina, 2017). Even sequential bilinguals, who may initially show delays in understanding macrostructure relative to simultaneous bilinguals or monolinguals, catch up quickly because this component of narrative comprehension is a skill at the cognitive-linguistic interface that can be transferred from the first language (Paradis et al., 2011; Pearson, 2002).

Weaker microstructure comprehension by bilingual children can be attributed to lower exposure to each of their two languages (Hoff et al., 2012; Thordardottir & Brandeker, 2013), which leads simultaneous and sequential bilingual children to score lower than monolingual peers on receptive vocabulary (Bialystok, Luk, Peets, & Yang, 2010; Morales, Calvo, & Bialystok, 2013) and expressive vocabulary (Nicoladis, 2003; Oller, 2005; Oller & Eilers, 2002), for one and frequently both languages.

Whether bilingual children show lower sentence comprehension than monolinguals is less clear. Haman and colleagues (2017) showed that bilingual Polish-English children scored lower than monolingual Polish children on sentence comprehension, as measured by an unpublished version of the TROG test; and lower on sentence production, as measured in a sentence repetition task. On the other hand, several studies have shown that the syntactic development of bilingual children does not lag behind that of monolingual children (Thordardottir, Rothenberg, Rivard, & Naves, 2009), and that bilingual children can outperform monolinguals in sentence comprehension in the presence of interference (e.g. Filippi, Morris, Richardson, Bright, Thomas, Karmiloff-Smith, & Marian, 2015).

Bilingual children appear to develop primary language (L1) and secondary language (L2) abilities more slowly than monolingual peers (Farnia & Geva, 2011) during the preschool years (Hammer, Lawrence, Rodriguez, Davison, & Miccio, 2011) and school years (Bialystok et al., 2010), even after three consecutive years of exposure to both languages (Thordardottir, 2011). This has caused concern among many about the academic outcomes of bilingual children, especially since a "vicious cycle" can occur in which children's poor vocabulary knowledge and sentence comprehension limit their full participation in the curriculum and therefore hinder further development of language skills (McWilliam, 1998).

Just few studies have examined the role of receptive vocabulary and sentence comprehension in narrative competence by bilingual children. Studies have reported correlations between vocabulary measures and narrative production in bilingual and monolingual pre-schoolers (Korecky-Kröll et. al., 2019; Uccelli & Páez, 2007). However, we are unaware of studies analyzing, in bilinguals, the relationship of narrative comprehension of pictured stories to the linguistic skills of vocabulary and sentence comprehension.

The few relevant studies suggest the possibility that language skills may contribute differently to L1 and L2 narrative comprehension. Uccelli and Páez (2007) found that receptive vocabulary was positively associated with narrative comprehension in both languages and concluded that the two types of understanding mutually reinforce each other, although receptive vocabulary develops more slowly than narrative comprehension. in bilingual children. At the same time, those authors failed to find associations for either type of comprehension between the L1 and L2.

1.4 The current study

The current study is aimed to extend the findings described above by investigating the relationships of receptive vocabulary and sentence comprehension to narrative comprehension of pictured stories by Croatian (L1)-Italian (L2) bilingual preschool children, by controlling for the influence of age, SES and bilingual exposure. This study attempted to fill several research gaps about language development in bilingual preschool children, including poor understanding of the contributions of receptive vocabulary and sentence comprehension to narrative comprehension before the start of formal education. We also wanted to investigate whether these two factors contribute differently to comprehension of narratives depending on the language. Our focus was on narrative comprehension rather than narrative production.

We hypothesized that the children would show lower receptive vocabulary and sentence comprehension in the L2 than in L1. This idea is based on literature showing poorer language skills among bilingual children who are not exposed to the two languages from birth (cf. Bonifacci et al., 2018). We further hypothesized that both receptive vocabulary and sentence comprehension would contribute to narrative comprehension, such that lower levels of the first two would be associated with lower comprehension. This idea is based on studies of linguistic skills in the development of narrative competence in mono- and bilingual speakers (Bishop & Adams, 1992; Uccelli & Páez, 2007). Finally, the lack of consensus in the literature prevented us from formulating a hypothesis about whether receptive vocabulary and sentence comprehension would contribute differently to L1 and L2 narrative comprehension.

2. Method

2.1 Participants

Participants in this study were bilingual speakers of Croatian (L1) and Italian (L2), members of the Italian-Croatian bilingual community in areas near the Italian border. Bilingualism in the areas of Istria and Rijeka is maintained and encouraged: bilingualism is official policy, and kindergartens and schools offer programs in Italian. Italian culture is readily accessible, and economic ties to Italy are strong. Participants were 30 preschool children 5 and 6 years old (mean age, 73 months, *SD* = 5 months, range = 61–84 months) who had been attending Italian kindergarten for at least 18 months prior to the study. In Croatia, children start to attend school after the age of 6;0, but children closer to 6;0 are encouraged to wait an additional year. Parents reported that their children had been exposed to both languages for more than two consecutive years. None of the participants were simultaneous bilinguals, and the mean period of exposure to Italian was 26 months. Children participated in the study during the second half of their final year of preschool education.

Croatian belongs to the family of South-Slavic languages. It is a fusional language in which bound grammatical morphemes (endings) usually express three grammatical categories (case, number and gender for nouns and adjectives; person, number and tense for verbs). The canonical word order is SVO, but, due to morphological features, word order is relatively free. Verbs are subdivided into nine conjugational classes (Jelaska, 2005) according to differences in stem alternations. All this makes for a rather rich morphological system.

Italian is a Romance language. It is a fusional language in which bound grammatical morphemes (endings) express two or three grammatical categories (number and gender for nouns; person, number and tense for verbs; number and gender for articles). Italian nouns do not have cases but do have plural forms. A classification by D'Achille and Thornton (2003) distinguishes six classes based on plural formation. Additionally, there are both four indefinite and seven definite articles. There are three regular conjugational classes for verbs.

The bilinguals in the present study live in the county of Istria in northwestern Croatia. The most widespread variant of the Italophone repertoire in Istria is Istrovenetian, a subvariety of the Venetian dialect (ISO 639–3 code: VEC), spoken in Italy. Istrovenetian is a community language (Poropat Jeletić, 2015), so children acquire it first and they hear it in their everyday life outside kindergartens, as well as in institutions. Children receive little exposure to standard Italian before formal schooling, when they have the option to be educated in Italian; their preschool exposure is limited to the media and stories in Italian that may be read to them.

Venetian as well as Istrovenetian are considered to be distant from standard Italian because of linguistic and historical specificities – in fact, Venetian is sometimes considered a separate language – but the distance is not more than what several other dialects in Italy show (Ferguson, 2013). The Croatian variant spoken in the subjects' part of Istria is the southwestern Istrian dialect. It is the main variant spoken in the community, and although standard Croatian is more present than standard Italian, children are still primarily exposed to it through media and schooling, while dialect is used in most other situations. The fact that participants speak dialects as their L1 and L2 may have influenced their performance on standardized tests, but our results should retain internal validity because all participants belonged to the same bilingual community and shared similar linguistic experiences.

2.2 Materials

2.2.1 *Narrative comprehension*

Narrative comprehension was assessed using elicitation narrative material of the Multilingual Assessment Instrument for Narratives (MAIN) (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012, 2015) adapted for Croatian (Hržica & Kuvač Kraljević, 2012) and for Italian (Roch & Levorato, 2012). MAIN was designed to assess narrative skills in children who acquire one or more languages from an early age, and it evaluates both comprehension and production of narratives. It can assess two languages in the participant since it contains four parallel stories, two for telling and two for retelling.

Two MAIN forms for telling were used, 'Baby Goats' and 'Baby Birds'. Picture material was presented according to the computer presentation model of MAIN. The child clicked on his or her choice of square, which initiated a PowerPoint® presentation. In fact, the selection process was simulated: the examiner had preselected the story. First, the child viewed the entire set of six pictures in the middle of the screen, after which he or she pressed a key on the keyboard to reveal each pair of pictures. During the initial viewing and telling of the story, only the participant could see the computer screen. After telling the story, the child answered a set of comprehension questions while looking at the sequence of six pictures.

Participants were divided into two groups, one of which told the 'Baby Goats' story in Croatian and 'Baby Birds' in Italian, while the other told 'Baby Birds' in Croatian and 'Baby Goats' in Italian.

Nine comprehension questions focusing on macrostrusture components and internal state terms were used within the Croatian and Italian version of MAIN. Three questions asked about the goals of each of the three episodes. Three questions focused on eliciting information about the internal state (IS) of a character (e.g. how was the character feeling?) during a particular event or reaction. Each of the questions also had a follow-up question that was asked only if a child did not explain his or her first answer. In the final version of MAIN, there is also a question number 10, tapping theory of mind/inferencing, asking whom the mother bird/goat loved more, cat/fox of dog/bird. This question was not included into this study. As the last question is different from the other 9 questions, we believe that the task performance is quite reliable without it.

Comprehension questions were introduced by informing the participant that he/she would be asked some questions about the story. An initial warm-up question, "Did you like the story?", was asked, followed by the real questions in the order shown in Table 1. While asking questions, the investigator pointed to specific pictures of the story.

Questions 1, 4 and 7 (goal-oriented questions) were scored by 1 point if answered appropriately. Questions 2, 5 and 8 (IS-oriented questions) were scored with 1 point if IS was appropriate to the story. When participants provided an appropriate explanation of their original answer, the answer was scored with 1 additional point. Questions 3, 6 and 9, if asked, were worth 1 point each.

Table 1. Type of questions used for the assessment of narrative comprehension

Q0. Warm-up question	
Q1. Episode 1 – goal	Episode 1
Q2. Episode 1 – IS as initiating event	
Q3. Episode 1 – IS (asked only if no explanation of Q2 answer)	
Q4. Episode 2 – goal	Episode 2
Q5. Episode 2 – IS as reaction	
Q6. Episode 2 – IS (asked only if no explanation of Q5 answer)	
Q7. Episode 3 – goal	Episode 3
Q8. Episode 3 – IS as reaction	
Q9. Episode 3 – IS (asked only if no explanation of Q8 answer)	

IS = internal state.

Comprehension was scored by a bilingual speaker of Croatian and Italian who had been trained in language research methodology and in coding of narrative comprehension. As a validation of this scoring, the comprehension of 10 participants in each language was independently scored by two experienced researchers (one for each language). The results showed excellent intra-class correlation (ICC = .98, 90% confidence interval .94–.99).

2.2.2 *Receptive vocabulary test*

Participants were tested using the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997), a receptive vocabulary test in which the participant is asked to choose one of four pictures upon hearing the target word. The two PPVT versions in the present study were adapted and standardized for Croatian (Dunn & Dunn, Kovačević, Padovan, Hržica, Kuvač Kraljević, Mustapić, Dobravac, & Palmović, 2010 – reliability: alpha 0.956) or Italian (Dunn & Dunn, Stella, Pizzoli, & Tressoldi, 2000). These adapted versions of the PPVT keep the same procedure as the original but change the lexical material (order of words, exclusion/inclusion of words). Since the PPVT versions used in this study were standardized, our results on receptive vocabulary could be compared to normative data, allowing objective assessment of skills in Croatian and Italian.

PPVT in both languages was scored by a trained bilingual speaker of Croatian and Italian, under the supervision of researchers experienced in language testing.

2.2.3 *Sentence comprehension*

Children were assessed using the Test for Reception of Grammar (TROG-2) (Bishop, 2003), which assesses understanding of a range of grammatical structures. TROG-2 is a receptive test in which the participant is asked to choose one of four pictures upon hearing the target sentence. Sentences are grouped in 20 blocks of four sentences each. Testing finishes when the participant makes at least one error in five consecutive blocks. The two versions of TROG-2 in the present study were adapted for Italian (Bishop, Suraniti, Ferri, & Neri, 2009) or for Croatian (Bishop, Kuvač Kraljević, Hržica, Kovačević, & Kologranić Belić, 2013 – reliability: alpha 0.956). These adapted versions keep the same procedure as the original but change the lexical material and types of sentences used.

Since the TROG-2 versions used in this study were standardized, our results on sentence comprehension could be compared to normative data, allowing objective assessment of skills in Croatian and Italian. The TROG-2 in both languages was scored by a trained bilingual speaker of Croatian and Italian, under the supervision of researchers experienced in language testing.

2.2.4 *Parental reporting*

Data on background measures were provided by parents, who filled out a 40-item questionnaire divided into four sections (adapted from Roch & Florit, 2013) only two of which were used in the analysis. The section on language status asked about languages spoken by family members, the languages that the children spoke, the frequency with which each language was used by children and their families in different contexts, as well as information about how and when each of the languages had been acquired. In particular, we used the following variables to assess participants' language:

a. *Age at onset of exposure to the L2:* parents were asked to indicate at what age their child was exposed for the first time to Italian.
b. *Daily amount of input from each language:* parents were asked to estimate where, with whom, and for how long their child spent time on an average day in a week, and which language(s) each person used when addressing the child (based on a 10-point scale) (input). In addition, we asked parents to estimate which language(s) the child used to answer that person, using a 10-point scale (output).
c. *Socioeconomic status:* in this section parents were asked to provide their highest educational level, years of education, and annual family income. We measured socioeconomic status (SES) as a composite score based on both parents' education level and annual family income. Data on both variables were collected from the questionnaire. Parental educational levels were classified into five categories: 1 = Middle school degree, 2 = High school degree, 3 = Bachelor's degree, 4 = Master's degree, and 5 = Post-graduate degree. Household income was coded based on a 5-point scale: 1 = far below the national mean income, 2 = below the national mean income, 3 = national mean income, 4 = above the national mean income, 5 = far above the national mean income.

2.3 Statistical analyses

Differences in receptive vocabulary, sentence comprehension and narrative comprehension between L1 and L2 were assessed for significance using paired *t* tests and after adjusting the standard p value by Bonferroni's correction for multiple comparisons (.05 / number of comparisons). Effect sizes were reported for each *t* test, and an effect size $d = 0.2$ was considered small; 0.5, medium; and 0.8, large. Thus, two means were considered to show a trivial difference (even if statistically significant) if they differed by less than 0.2 standard deviations.

Potential relationships between participant characteristics and the three language comprehension variables (receptive vocabulary, sentence comprehension, narrative comprehension), as well as relationships among the three comprehension variables, were assessed using Pearson's r. Two-tailed significance was determined after adopting Bonferroni's correction for multiple comparisons.

Finally, multivariate hierarchical regressions were performed to analyze the contribution of vocabulary and sentence comprehension to narrative comprehension in the L1 and L2. Adjusted R^2 and change in R^2 were evaluated in order to control for multiple predictors.

3. Results

3.1 Descriptive statistics

The first set of results describes the background characteristics and language skills of the participants.

3.1.1 *Socioeconomic status*

We calculated a composite SES score including the education level of both parents and family income, since income and education levels were positively related to one another ($0.40 < r < 0.51$ $p = 0.001$). The composite SES variable was constructed using principal component analysis. SES of our participants was medium to low ($M = 1.5$, $SD = 2.7$): 58% of families had low SES, 35% had medium SES, and only 7% had high SES (Table 2). The education level was low for 64% of mothers, while the remaining 36% had a university education. The corresponding percentages for fathers were similar (72% and 28%).

Table 2. Participant characteristics in socioeconomic status (SES)

Category	Family income	Education level	Mother, n	Father, n	Composite SES
0 – very low	33.3	Secondary school	0	4.5	26.9
1 – low	20.8	High school	64	68.2	30.8
2 – medium	20.8	Bachelor	4	18.2	35
3 – high	8.3	Master	28	9.1	3.5
4 – very high	16.7	Postgraduate	4	0	3.5

3.1.2 *Language status*

Two broad measures of exposure to each language were determined from parental reports. The first measure was the age at onset of bilingual exposure, namely how old the child was when he or she began to be exposed to the L2. The mean age at onset of bilingual exposure was during the third year of life ($M = 32$ months, $SD = 7$ months), indicating that the participants may be defined as sequential bilingual speakers.

The second measure was the children's daily language exposure, in terms of input (what they heard or saw) and output (what they themselves produced) during interaction with persons frequently in contact with the child in everyday contexts, including schoolmates, teachers, friends, family, grandparents, brothers and sisters.

Figure 1 shows that on a typical day, participants received a relatively balanced input in L1 and L2. There was a strong correspondence between the language provided to the children and the language in which the children responded.

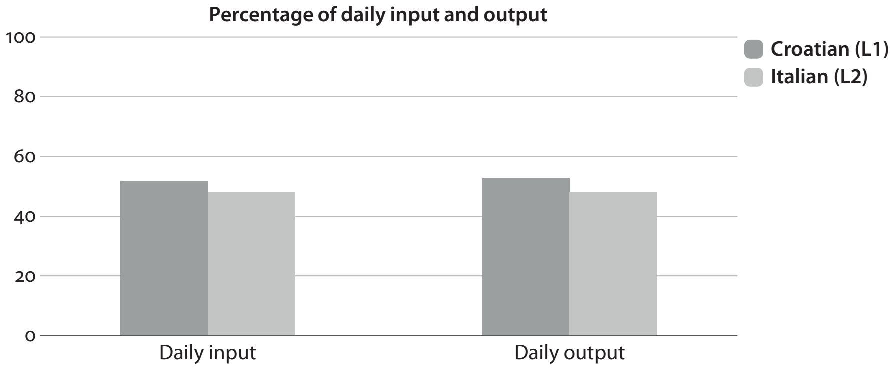

Figure 1. Daily language input and output of participants, by L1 or L2. Each type of input in each language is shown as a percentage of the total daily exposure

3.1.3 *Language skills*

Table 3 presents descriptive statistics of the language skills for the 30 participants.

Table 3. Language skills of participants

Test		Mean (SD)	Paired t test
PPVT (receptive vocabulary) (M = 100, SD = 15)	L1 (Croatian)	95.7 (16.3)	$t(29) = 5.3$, $p < .001$, $d = 1.63$
	L2 (Italian)	70.5 (14.6)	
TROG (sentence comprehension) (M = 100, SD = 15)	L1 (Croatian)	91.9 (11.9)	$t(29) = 2.4$, $p < .05$, $d = .399$
	L2 (Italian)	86.6 (17.8)	
Narrative comprehension (Max = 9)	L1 (Croatian)	6.5 (2.1) 72%	$t(29) = 2.8$, $p < .001$, $d = .604$
	L2 (Italian)	5.1 (2.8) 57%	

On tests of general language abilities, participants scored significantly better in the L1 than in L2, and effect sizes were large for receptive vocabulary and medium-low for sentence comprehension. Participants showed monolingual-like performance in receptive vocabulary and sentence comprehension in the L1, but they scored two standard deviations below the mean in the L2 receptive vocabulary test and one standard deviation below the mean in the L2 sentence comprehension test. Narrative comprehension was also significantly better in L1 that in L2 and the effect size was medium-low. These results are in line with those reported in previous studies evaluating bilingual children 5–7 years old using the same narrative comprehension task. Those studies reported results higher from 10% to 16% in L1, when compared to L2 (Bohnacker, 2016; Roch et al., 2016; Rodina, 2017).

To assess narrative comprehension in greater detail, we compared the children's performance on comprehension questions about implicit information (characters' goals and internal states) between L1 and L2 (Table 4).

Table 4. Mean scores and correct response rates on narrative comprehension questions about characters' goals and internal states

Type of comprehension question	L1 (Croatian)	L2 (Italian)	t test
Goals (score, Max = 3)	2.6 (.77) 87%	2.1 (1.1) 70%	$t\ (29) = 2.6\ p < .01\ d = .52$
Internal states (score, Max = 6)	3.9 (1.7) 65%	3.0 (1.8) 50%	$t\ (29) = 2.6\ p < .01\ d = .51$

As with the individual language comprehension skills, the children performed better in the L1 than L2 for comprehension of goals and internal states. The effect size was medium for goals and internal states, reflecting the difference of 15 percentage points in correct response rates between the two languages.

Regardless of language, correct responses to questions about goals were 20% higher than for questions about internal states. This means that the children were much better at understanding goals than at understanding the motives underlying those goals or the actions designed to achieve them. These results are similar to those reported by Bohnacker (2016) and Rodina (2017) among bilingual children 5–7 years old on the same narrative comprehension task.

3.2 Relationships between participant characteristics and language skills

We explored potential correlations between participant characteristics (age, SES, age at onset of L2 exposure, daily amount of exposure to L1 and L2) and language skills (receptive vocabulary, sentence comprehension, narrative comprehension). We explored such correlations separately in the L1 and L2. Standardized scores on the language comprehension tests were used in order to control for age. No significant correlations were observed. This likely reflects the low variation in participant characteristics but relatively high variation in language skills. For this reason, background characteristics were excluded from all subsequent analyses aimed at investigating the role of vocabulary and sentence comprehension in narrative comprehension.

3.3 The role of receptive vocabulary and sentence comprehension in narrative comprehension

Next we investigated to what extent individual differences in narrative comprehension are determined by differences in linguistic comprehension (vocabulary and sentence comprehension), within each language and across the two languages. Tables 5a and 5b report the correlations among the three measures in each language, while Table 5c reports the correlations between the two languages.

Table 5a. Correlations among receptive vocabulary, sentence comprehension and narrative comprehension in the L1 (Croatian)

	PPVT	TROG	Narrative comprehension
PPVT	1	.380*	.487**
TROG		1	.492**
Narrative comprehension			1

Note. * $p < .05$; ** $p < .01$

Table 5b. Correlations among receptive vocabulary, sentence comprehension and narrative comprehension in the L2 (Italian)

	PPVT	TROG	Narrative comprehension
PPVT	1	.728**	.293
TROG		1	.44**
Narrative comprehension			1

*Note.*** $p < .01$

The results in Table 5a show that the three levels of language comprehension, from word to text comprehension, correlate with one other. The results in Table 5b show that sentence comprehension correlates with both receptive vocabulary and narrative comprehension in the weaker language, while there is no significant relationship between receptive vocabulary and narrative comprehension.

Table 5c. Correlations among receptive vocabulary, sentence comprehension and narrative comprehension between the L1 and L2

	PPVT L2	TROG L2	Narrative comprehension L2
PPVT L1	−.413**	.093	.200
TROG L1	.301	.722**	.262
Narrative comprehension L1	−.196	.189	.391*

*Note.** $p < .05$; ** $p < .01$

The results in Table 5c show that each linguistic comprehension ability in one language correlates with the same ability in the other language, suggesting interdependence between the two languages. Surprisingly, while this relationship is positive for L1-L2 sentence and narrative comprehension, it is negative for L1-L2 receptive vocabulary, suggesting competition between vocabulary knowledge in the two languages. The cross-linguistic interdependence is weak: the three comprehension measures in the L1 do not correlate, respectively, with the other in the L2. In other words, the three levels of language comprehension correlate with one another in each of the two languages, but not between the two languages. Cross-linguistic correlations emerged only for the same measure in the two languages.

Multivariate regression was used to analyze the contribution of receptive vocabulary and sentence comprehension to narrative comprehension. These regression analyses were performed separately for each of the two languages using the same potential predictors. In the first step, receptive vocabulary and sentence comprehension in the same language of the dependent variable were inserted. In the second step, receptive vocabulary and sentence comprehension in the other language than the dependent variable were inserted. Tables 6a and 6b illustrate the results of narrative comprehension in the L1 and L2, respectively. Adjusted R^2 values were used because of the small sample and multiple predictors.

The two regression analyses indicated that the variance in narrative comprehension is much greater in the L1 than in the L2. They further indicated that receptive vocabulary and sentence comprehension account for 29.7% and 13.4% of variance in narrative comprehension in the L1 and L2, respectively. The two language comprehension measures in one language do not account for variance in narrative comprehension in the other language over and above the variance in

Table 6a. Multivariate regression to predict narrative comprehension in the L1

		R^2change		**B**	**SE B**	**β**
Step 1	PPVT L1	$.354^{\circ\circ}$				
	TROG L1	Adj R^2 .306	PPVT L1	.047	.023	.351*
			TROG L1	.069	.033	.374*
Step 2	PPVT L1	$.040^{\circ\circ\circ}$				
	TROG L1	Adj R^2 .297	PPVT L1	.026	.031	.194
	PPVT L2		TROG L1	.094	.051	.514
	TROG L2		PPVT L2	−.038	.051	−.254
			TROG L2	−.001	.049	−.007

Note. Adjusted R^2 = .297 $F(5, 29) = 3.16$, $p < .05$; ° F change (1, 24) = .032, $p = .943$; °° F change (2, 26) = 7.24, $p < .001$; °°° F change (2, 24) = .77, $p = .71$; * $p < .05$

Table 6b. Multivariate regression to predict narrative comprehension in the L2

		R^2change		B	SE B	β
Step 1	PPVT L2	.197°°				
	TROG L2	Adj R^2 .137	PPVT L2	.047	.023	.351
			TROG L2	.069	.033	.374*
Step 2	PPVT L2	.057°°°				
	TROG L2	Adj R^2 .134	PPVT L2	.026	.031	.194
	PPVT L1		TROG L2	.094	.051	.514
	TROG L1		PPVT L1	−.038	.051	−.254
			TROG L1	−.001	.049	−.007

Note. Adjusted $R^2 = .134$, $F(5, 29) = 3.2$, $p < .05$; ° F change $(1, 28) = .067$, $p = .797$; °° F change $(2, 26) = 3.14$, $p < .05$; °°° F change $(2, 24) = .946$, $p = .402$; * $p < .05$

the same language. In other words, narrative comprehension in one language is predicted by the two language comprehension measures within the same language, but not in the other language.

The regression analyses also indicated that while both receptive vocabulary and sentence comprehension contribute significantly to narrative comprehension in the L1, only sentence comprehension contributes significantly in the L2.

4. Discussion

The current study aimed to analyze narrative comprehension of pictured stories in the L1 and L2 of bilinguals 5–7 years old, in terms of their receptive vocabulary and sentence comprehension. To the best of our knowledge this is the first study of narrative comprehension in Croatian-Italian bilinguals. In contrast to many studies on bilingual migrants or heritage language speakers, the participants of our study live in an area where bilingualism is institutionalized.

The children performed better in their L1, and the effect size was high for vocabulary and medium-low for sentence and narrative comprehension. Linguistic abilities contributed in different ways to L1 and L2 narrative comprehension. In the L1, receptive vocabulary and sentence comprehension together explained 29% of variance in narrative comprehension, but in the L2, sentence comprehension alone predicted 13.4% of narrative comprehension, whereas receptive vocabulary predicted much less. Linguistic skills in the two languages did not substantially interact: receptive vocabulary and sentence comprehension in one language did not correlate with narrative comprehension in the other.

This study makes at least two contributions to existing knowledge about narrative comprehension in bilingual children. First, we provide evidence that the difference in narrative comprehension between L1 and L2 is paralleled by differences in receptive vocabulary and sentence comprehension. Second, we provide new insights into how receptive vocabulary and sentence comprehension may contribute to narrative comprehension in general, and we provide evidence that these contributions depend, in bilinguals, on whether the language is L1 or L2.

4.1 Narrative comprehension in L1 and L2

Our participants showed better narrative comprehension in their L1 than L2, as reported for other bilinguals of similar age (Bohnacker, 2016; Roch et al. 2016; Rodina, 2017). Independently of the language, our participants were more accurate at answering questions about characters' goals than questions about characters' mental states, as reported by Bohnacker (2016), Rodina (2017) and Lindgren (2018), using the same narrative comprehension task. This may reflect that children in this age range struggle with the comprehension of motivations and emotions. It appears, therefore, that MAIN narrative comprehension task is sensitive enough to capture differences in the comprehension of specific narrative components by preschool children. Age did not significantly affect narrative comprehension in our sample, in contrast to a previous study involving bilinguals 5–7 years old and the same task (Roch et al., 2016). Future studies should address the age-dependence of narrative comprehension.

The L1-L2 advantage in linguistic skills and in narrative comprehension in our sample doubtless reflects their status as sequential, rather than simultaneous, bilinguals. The participants showed balanced current exposure to both languages, and they varied relatively little in SES and age at onset of bilingual exposure, which probably explains why these background variables did not significantly influence narrative comprehension. For this reason, we feel confident that our results are representative of children showing a typical sequential bilingual linguistic trajectory. Indeed, our sample showed substantially lower receptive vocabulary than their monolingual peers in both languages (Uccelli & Páez, 2007). The delay for sentence and narrative comprehension is smaller but still the performance in L2 is significantly lower than in L1.

4.2 Associations between language skills and languages

In our sample, all three levels of language comprehension correlated with one another in the L1, but only two of the three levels correlated with each other in the L2. This suggests that linguistic comprehension shows a monolingual-like pattern of relationships from word to narrative comprehension in L1, and that the second language in bilinguals develops differently from the first language. In the first years of bilingual development, the use of two languages is cognitively demanding and cognitive mechanisms might not be sufficiently efficient at distributing resources between the two languages adequately (Bialystok & Craik, 2010). As children gain experience in both languages, they may be able to deploy cognitive resources efficiently between the two languages. Future studies should investigate cognitive allocation in bilingual speakers.

Our results suggest a degree of interdependence between L1 and L2: each language comprehension measure in one language correlated with the same measure in the other language. Each comprehension level in one language however, correlated weakly with the other levels in the other language. Other studies have similarly reported strong correlation between vocabulary and sentence comprehension within one language but weak relationships between the two languages (e.g. Marchman et al., 2004; Uccelli & Páez, 2007). Our results, together with the literature, highlight the complexity of the relationships between L1 and L2 in bilinguals, which should be further understood, especially in preschool children, as a way to ensure school readiness.

For instance, future work may wish to follow up on our intriguing finding of a positive correlation of sentence and narrative comprehension between the two languages, suggesting linguistic interdependence in these domains, but a negative correlation of receptive vocabulary, suggesting vocabulary competition. Our data suggest that in bilinguals 5–7 years old, transfer of sentence and narrative comprehension can occur without transfer of receptive vocabulary. This may be because vocabulary is more strongly influenced by experience than the sentence and narrative comprehension are. A second potential explanation is that vocabulary transfers little or not at all between Italian and Croatian because the languages are too different: lexical transfer is more extensive between closely related languages that share many cognates (e.g. Lindgren & Bohnacker, 2020). Future studies should address these and other aspects of the relationships between L1 and L2 in bilinguals.

Our results about the contributions of vocabulary and sentence comprehension to narrative comprehension in L1 and L2 align with findings from previous work on the role of language skills in oral narrative comprehension in pre-schoolers (e.g. Florit et al., 2013). In this way, our study extends those findings to pictured narratives. It appears that the way in which linguistic skills contribute to the narrative comprehension of pictured stories in the L1 of bilinguals is similar to the way in which they contribute to oral narrative comprehension by monolingual speakers of the same age (Lepola et al., 2012).

The present study extends the literature by showing that linguistic skills contribute differently to comprehension in the L1 or L2. For instance, in the L1, both receptive vocabulary and sentence comprehension contribute strongly to narrative comprehension; but in the L2, sentence comprehension plays an important role in narrative comprehension, while vocabulary contributes weakly. It could be argued, therefore, that bilinguals rely on different mechanisms during narrative comprehension in their two languages. Indeed, the two linguistic skills in our study accounted for 29% of variance in narrative comprehension in the L1 but only 13% in the L2.

The weaker contribution of linguistic skills in the L2 appears to depend largely on poorer vocabulary. It may be that during L2 narrative comprehension, children tend to rely less on linguistic skills, probably because they are weak, especially vocabulary. Instead, they may rely on cognitive skills that we did not evaluate here, such as inferential abilities, comprehension monitoring and knowledge of the story structure, all of which can predict narrative comprehension (Cain & Oakhill, 2007). Inferential skills, for example, are crucial for oral narrative comprehension by preschool monolingual speakers (Florit et al. 2014). Inferential skills may be particularly important for bilingual speakers trying to understand pictured narratives. Future studies should analyze cognitive as well as linguistic factors in predicting bilingual narrative comprehension.

5. Conclusion

In our study of Croatian-Italian bilingual preschool children, receptive vocabulary and sentence comprehension emerged as equally important components of narrative comprehension in the L1, exhibiting a monolingual-like pattern of relations. However, these linguistic skills, particularly vocabulary, were much less important for narrative comprehension in the weaker L2. These findings highlight the need to investigate broader linguistic comprehension in bilinguals before they begin formal education and develop reading comprehension. Future studies should also investigate the contribution of non-linguistic skills to narrative comprehension, which also depends on higher-level cognitive skills such as inference-making, comprehension monitoring, knowledge of story structure and working memory. These cognitive components may be even more important for narrative comprehension in bilinguals than monolinguals.

Funding

Funded by: European Commission Seventh Framework Programme.
Funded by: Croatian Science Foundation.
Award ID: UIP-2017-05-6603.
Statement: This work has been supported by the European Commission Seventh Framework Programme for the project *Language Dominance of Bilingual Speakers Perceived as Balanced* (LADOBI, PISCRTJQ4Y, Piscopia Marie Curie Action (COFUND), Framework 7 (FP7)), and by the Croatian Science Foundation for the project *Multilevel approach to spoken discourse in language development* (UIP-2017-05-6603).

References

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, 19(1), 19–23. https://doi.org/10.1177/0963721409358571

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bishop, D. (2003). *Test for reception of grammar, Version 2*. London: The Psychological Corporation.

Bishop, D., & Adams, C. (1992). Comprehension problems in children with specific language impairment: literal and inferential meaning. *Journal of Speech and Hearing Research*, 35(1), 119–129. https://doi.org/10.1044/jshr.3501.119

Bishop, D., Kuvač Kraljević, J., Hržica, G., Kovačević, M., & Kologranić Belić, L. (2013). *TROG-HR Test razumijevanja gramatike*. Zagreb: Naklada Slap.

Bishop, D., Suraniti, S., Ferri, R., & Neri, V. (2009). *Test for reception of grammar, version 2 TROG-2*. Florence: Edizioni, Italy: Giunti OS.

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

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Bonifacci, P., Barbieri, M., Tommasini, M., & Roch, M. (2018). In few words: linguistic gap but adequate narrative structure in preschool bilingual children. *Journal of Child Language*, 45(1), 120–147. https://doi.org/10.1017/S0305000917000149

Cain, K., & Oakhill, J. (2009). Reading comprehension development from 8 to 14 years: The contribution of component skills and processes. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp. 143–175). New York, NY: The Guildford Press.

Cain, K., & Oakhill, J. V. (2007). Reading comprehension difficulties: Correlates, causes and consequences. In K. Cain & J. V. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 3–40). New York, NY: Guilford Press.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96(1), 31–42. https://doi.org/10.1037/0022-0663.96.1.31

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference making ability, and their relation to knowledge. *Memory and Cognition*, 29(6), 850–859. https://doi.org/10.3758/BF03196414

Cleave, P. L., Girolametto, L. E., Chen, X., & Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning children with specific language impairment. *Journal of Communication Disorders*, 43(6), 511–522. https://doi.org/10.1016/j.jcomdis.2010.05.005

D'Achille, P., & Thornton, A. M. (2003). La flessione del nome dall'italiano antico all'italiano contemporaneo. In N. Maraschio & T. Poggi Salani (Ed.), *Italia Linguistica Anno Mille – Italia Linguistica Anno Duemila. Atti del XXXIV congresso internazionale di studi della SLI* (pp. 211–230). Roma: Bulzoni.

De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language & Communication Disorders*, 46(6), 613–627. https://doi.org/10.1111/j.1460-6984.2011.00049.x

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-PPVT – 3rd edition*. Minneapolis, MN: American Guidance Service.

Dunn, L. M., & Dunn, L. M., Kovačević, M., Padovan, N., Hržica, G., Kuvač Kraljević, J., Mustapić, M., Dobravac, G., & Palmović, M. (2010). *PPVT-III-HR Peabody slikovni test rječnika*. Zagreb: Naklada Slap.

Dunn, L. M., & Dunn, L. M., Stella, G., Pizzoli, C., & Tressoldi, P. (2000). *PPVT-revised*. Torino: Omega Edition.

Ferguson, R. (2013). Venetian language. In E. Dursteler (Ed.) *A companion to Venetian history, 1400–1797* (pp. 929–957). Leiden: Brill. https://doi.org/10.1163/9789004252523_027

Filippi, R., Morris, J., Richardson, F. M., Bright, P., Thomas, M. S., Karmiloff-Smith, A., & Marian, V. (2015). Bilingual children show an advantage in controlling verbal interference during spoken language comprehension. *Bilingualism*, 18(3), 490–501. https://doi.org/10.1017/S1366728914000686

Florit, E., Roch, M., Altoe, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology*, 27(4), 935–951. https://doi.org/10.1348/026151008X397189

Florit, E., Roch, M., & Levorato, M. C. (2011). Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. *Discourse Processes*, 48(2), 119–138. https://doi.org/10.1080/0163853X.2010.494244

Florit, E., Roch, M., & Levorato, M. C. (2013). The relationship between listening comprehension of text and sentences in preschoolers: Specific or mediated by lower and higher level components? *Applied Psycholinguistics*, 34(2), 395–415. https://doi.org/10.1017/S0142716411000749

Florit, E., Roch, M., & Levorato, M. C. (2014). Listening text comprehension in preschoolers: A longitudinal study on the role of semantic components. *Reading and Writing*, 27(5), 793–817. https://doi.org/10.1007/s11145-013-9464-1

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics 56*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–269). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-011

Gagarina, N., Klop, D., Tsimpli, I., & Walters, J. (2016). Narrative abilities in bilingual children. *Applied Psycholinguistics*, 37(1), 11–17. https://doi.org/10.1017/S0142716415000399

Farnia, F., & Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied. Psycholinguistics*. 32(4), 711–738. https://doi.org/10.1017/S0142716411000038

Haman, E., Wodniecka, Z., Marecka, M., Szewczyk, J., Białecka-Pikul, M., Otwinowska, A., Mieszkowska, K., Łuniewska, M., Kołak, J., Miękisz, A., Kacprzak, A., Banasik, N., & Foryś-Nogala, M. (2017). How does L1 and L2 exposure impact L1 performance in bilingual children? Evidence from Polish-English migrants to the United Kingdom. *Frontiers in Psychology*, 8, 1444. https://doi.org/10.3389/fpsyg.2017.01444

Hammer, C. S., Lawrence, F., Rodriguez, B., Davison, M. D., & Miccio, A. W. (2011). Changes in language usage of Puerto Rican mothers and their children: Do gender and timing of exposure to English matter? *Applied Psycholinguistics*, 32(2), 275–297. https://doi.org/10.1017/S014271641000041X

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1–27. https://doi.org/10.1017/S0305000910000759

Hoff, E., & Core, C. (2015). What clinicians need to know about bilingual development. *Seminars in Speech and Language*, 36(2), 89–99. https://doi.org/10.1055/s-0035-1549104

Hrzica, G., & Kuvač Kraljević, J. (2012). MAIN Croatian adaptation. In N. Gagarina, D. Klop, S. Kunnari, K. Tantele, T. Välimaa, I. Balčiūnienė, U. Bohnacker, & J. Walters, MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics 56*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, 15(1), 58–74. https://doi.org/10.1017/S1366728911000538

Jelaska, Z. (2005). *Hrvatski kao drugi i strani jezik*. Zagreb: Hrvatska sveučilišna naklada.

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The index of narrative microstructure: A clinical tool for analyzing schoolage children's narrative performances. *American Journal of Speech-Language Pathology*, 15(2), 177–191. https://doi.org/10.1044/1058-0360(2006/017)

Kendeou, P., van den Broek, P., White, M., & Lynch, J. S. (2007). Comprehension in preschool and early elementary children: Skill development and strategy interventions. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 27–45). Mahwah, NJ: Lawrence Erlbaum Associates.

Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, 31(3), 259–272. https://doi.org/10.1111/j.1467-9817.2008.00370.x

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of decoding and oral language skills. *Journal of Educational Psychology*, 101(4), 765–778. https://doi.org/10.1037/a0015956

Kim, Y. S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology*, 141, 101–120. https://doi.org/10.1016/j.jecp.2015.08.003

Kintsch, W. (1994). The psychology of discourse processing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539–588). San Diego, CA: Academic Press.

Korecky-Kröll, K., Dobek, N., Blaschitz, V., Sommer-Lolei, S., Boniecki, M., Uzunkaya-Sharma, K., & Dressler, W. U. (2019). Vocabulary as a central link between phonological working memory and narrative competence: Evidence from monolingual and bilingual four-year-olds from different socioeconomic backgrounds. *Language and Speech*, 62(3), 546–569. https://doi.org/10.1177/0023830918796691

Lindgren, J. (2018). *Developing narrative competence: Swedish, Swedish-German and Swedish-Turkish children aged 4–6* (Studia Linguistica Upsaliensia 19). Uppsala: Acta Universitatis Upsaliensis.

Lindgren, J., & Bohnacker, U. (2020). Vocabulary development in closely-related languages: Age, word type and cognate facilitation effects in bilingual Swedish-German preschool children. *Linguistic Approaches to Bilingualism*, 10(5), 587–622. https://doi.org/10.1075/lab.18041.lin

Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4–6-year-old children. *Reading Research Quarterly*, 47(3), 259–282. https://doi.org/10.1002/rrq.020

Long, D. L., Oppy, B. J., & Seely, M. R. (1997). Individual differences in readers' sentence and text level representations. *Journal of Memory and Language*, 36(1), 129–145. https://doi.org/10.1006/jmla.1996.2485

Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: evidence from bilingual language learners. *Developmental Science*, 7(2), 212–224. https://doi.org/10.1111/j.1467-7687.2004.00340.x

McWilliam, N. (1998). *What's in a word: Vocabulary development in multilingual classrooms*. Stoke-on-Trent: Trentham.

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114(2), 187–202. https://doi.org/10.1016/j.jecp.2012.09.002

Nation, K., Clarke, P., Marshall, C., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language and Hearing Research*, 47(1), 199–211. https://doi.org/10.1044/1092-4388(2004/017)

Nicoladis, E. (2003). Cross-linguistic transfer in deverbal compounds of preschool bilingual children. *Bilingualism: Language and Cognition*, 6(1), 17–31. https://doi.org/10.1017/S1366728903001019

Oakhill, J., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornoldi, & J. Oakhill (Eds.), *Reading comprehension difficulties. Processes and intervention* (pp. 69–92). Mahwah, NJ: Lawrence Erlbaum Associates.

Oakhill, J. V., & Cain, K. (2007). Introduction to comprehension development. In K. Cain, & J. V. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41–73). New York, NY: Guilford Press.

Oller, D. K. (2005). The distributed characteristic in bilingual learning. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds), *ISB4: Proceedings of the 4th International Symposium on Bilingualism* (pp. 1744–1749). Somerville, MA: Cascadilla Press.

Oller, D. K., & Eilers, R. E. (2002). *Language and literacy in bilingual children*. Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853595721

Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development and disorders*. Baltimore, MD: Paul H. Brookes.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skills. In M. Snowling, & C. Hulme (Eds.), *The science of reading: A handbook*. (pp. 227–247). Oxford: Blackwell. https://doi.org/10.1002/9780470757642.ch13

Pesco, D., & Kay-Raining Bird, E. (2016). Perspectives on bilingual children's narratives elicited with the Multilingual Assessment Instrument for Narratives. *Applied Psycholinguistics*, 37(1), 1–9. https://doi.org/10.1017/S0142716415000387

Poropat Jeletić, N. (2015). Italian language in Istria: Status planning, corpus planning and acquisition planning. *Mediterranean Journal of Social Sciences*, 6(2), 385–392.

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37(1), 49–67. https://doi.org/10.1017/S0142716415000417

Roch, M., & Florit, E. (2013). Narratives in preschool bilingual children: The role of exposure. *Rivista di Psicolinguistica Applicata*, 13(2), 55–63.

Roch, M., & Levorato, C. (2012). MAIN Italian adaptation. In N. Gagarina, D. Klop, S. Kunnari, K. Tantele, T. Välimaa, I. Balčiūnienė, U. Bohnacker, & J. Walters, MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics, 56*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian-Russian children. *International Journal of Bilingualism*, 21(5), 617–635. https://doi.org/10.1177/1367006916643528

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research*, 95(5), 259–272. https://doi.org/10.1080/00220670209596600

Seigneuric, A., & Ehrlich, M. F. (2005). Contribution of working memory capacity to children's reading comprehension. A longitudinal investigation. *Reading and Writing*, 18(7–9), 617–656. https://doi.org/10.1007/s11145-005-2038-0

Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology*, 107(2), 321–331. https://doi.org/10.1037/a0037769

Simon-Cereijido, G., & Gutiérrez-Clellen, V. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, 30(2), 315–337. https://doi.org/10.1017/S0142716409090134

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing*, 4(3), 245–256. https://doi.org/10.1007/BF01027150

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, 15(4), 426–445. https://doi.org/10.1177/1367006911403202

Thordardottir, E., Rothenberg, A., Rivard, M. E., & Naves, R. (2009). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, 4(1), 1–21. https://doi.org/10.1080/14769670500215647

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, 46(1), 1–16. https://doi.org/10.1016/j.jcomdis.2012.08.002

Uccelli, P., & Paéz, M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associationsamong English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, 38(3), 225–236. https://doi.org/10.1044/0161-1461(2007/024)

Van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M. J., & Purgzles Lorch, E. (2005). Assessment of Comprehension abilities in young children. In S. G. Paris, & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum Associates.

Bilingual children's lexical and narrative comprehension in Dutch as the majority language

Elma Blom and Tessel Boerma
Utrecht University

Background. Some studies find that bilingual children perform below monolinguals on language measures, whereas other studies report no differences. For the purpose of this study, we investigated the Dutch lexical and narrative comprehension of bilingual Tarifit-Dutch and Turkish-Dutch children. We compared the bilingual and monolingual children's performance. Within the bilingual groups, we explored relationships with language input at home.

Methods. 114 children participated (38 Tarifit-Dutch, 31 Turkish-Dutch, 45 monolingual Dutch), most of whom were 5 or 6 years old at the first testing time (mean age = 5.71 years, standard deviation = 0.64). The children were tested three times with one year between each testing time. Lexical and narrative comprehension were investigated with the Dutch version of the Peabody Picture Vocabulary Test (PPVT; Schlichting, 2005) and the Dutch version of the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012), respectively. Both answers to listening comprehension questions and questions after story generation were analyzed.

Results. Bilingual children performed lower on lexical comprehension than monolingual children. Narrative comprehension showed few differences across the two groups: the monolinguals performed slightly better on the comprehension questions after story generation at time 1, but not at times 2 and 3. Between-group comparisons within the bilingual group showed no differences in lexical comprehension of the Turkish-Dutch and Tarifit-Dutch children. The Turkish-Dutch children performed better than the Tarifit-Dutch children on narrative comprehension after story generation (time 2). In the bilingual sample, most relations between input at home and lexical and narrative comprehension were not significant, except for the positive correlations between socioeconomic status (SES) and lexical comprehension (times 2 and 3) in the Turkish-Dutch sample and home language richness and listening comprehension in the Tarifit-Dutch sample (time 1).

https://doi.org/10.1075/sibil.61.07blo

Conclusions and implications. Results of the current study showed larger gaps between monolinguals and bilinguals for lexical comprehension compared to narrative comprehension, suggesting that bilingual children's comprehension of narratives in the majority language benefits more from transfer than lexical comprehension does. Previous observations regarding the impact of input at home might not hold across all bilingual groups. Different home practices and cultures may moderate the effects of language richness on narrative comprehension. The same is true for the impact of SES on bilingual children's lexical comprehension. Conclusions about bilingual language development cannot be generalized across language domains or across bilingual populations.

Keywords: narrative comprehension, lexical comprehension, socioeconomic status, linguistic home environment, Tarifit-Dutch, Turkish-Dutch

1. Introduction

Research on childhood bilingualism has produced results that are at first sight contradictory. Several studies find that bilingual children, when tested in one of their languages, lag behind their monolingual peers (Bialystok, Luk, Peets, & Yang, 2010; Gathercole & Hoff, 2007; Hoff, Core, Place, Rumiche, Señor, & Parra, 2012), whereas other research observes no differences in performance (De Houwer, Bornstein, & Putnik, 2013; Hipfner-Boucher, Milburn, Weitzman, Greenberg, Pelletier, Girolametto, 2015; Pearson, 2002; Pearson, Fernández, & Oller, 1993). The primary goal of the present study was to determine whether differences between bilingual and monolingual children in language outcomes are modulated by language domain. To this end, we compare bilingual and monolingual children's lexical and narrative comprehension in Dutch, which is the majority language in the context in which the study was conducted (the Netherlands). Data from both groups were collected during the first years of elementary school at three times with yearly intervals. Owing to accumulating exposure to Dutch at school, bilingual children with a migrant background may become Dutch-dominant over time (Extra, Aarts, Van der Avoid, Broeder, & Yağmur, 2001). We compare bilinguals and monolinguals at all three times. The secondary goal was to explore variation within the group of bilingual children by comparing two bilingual subgroups from respectively Turkish and Moroccan descent. First, we compare Dutch lexical and narrative comprehension across the two groups. Second, we investigate if input factors in the two bilingual subgroups are related to Dutch lexical and narrative comprehension.

The results of this study may have clinical and educational repercussions. Lexicon and narration are often evaluated as part of standardized language assessments to determine the presence of an innate language disorder. If lexical and narrative comprehension are influenced by bilingualism, lexical and narrative assessment instruments should be used with caution in bilingual contexts, because they could lead to overdiagnosis (Engel de Abreu, Baldassi, Puglesi, & Befi-Lopez, 2013). Furthermore, previous research has demonstrated that weak lexical and narrative abilities indicate a risk of impaired or delayed literacy development (Snow, Burns, & Griffin, 1998; Dickinson & Tabors, 2001; Oller & Pearson, 2002; August & Shanahan, 2006). Determining any delays in these areas is particularly important for bilingual children with low socioeconomic (SES) backgrounds, who face the risk of low academic achievement (OECD, 2018). The current study included bilingual children from low SES backgrounds raised in migrant families.

Before turning to literature on lexical and narrative development in relation to bilingualism, we discuss the role of the input. The reason for this is twofold. First, in the literature, differences in input have been put forward to explain why the language development of bilingual children may lag behind that of their monolingual peers. Second, input may affect bilingual children's lexical and narrative comprehension differently.

1.1 The role of the input

It is well-established that language input plays an integral role in children's language development (Tomasello, 2003). The foundational role of language input may explain why bilingual children's language development can lag behind: bilinguals are likely to receive less input in either language compared to their monolingual peers because their total amount of input is divided over two languages (Gathercole & Thomas, 2009; Quiroz, Snow, & Zhao, 2010; MacLeod, Fabiano-Smith, Boegner-Pagé, & Fontolliet, 2013; Unsworth, 2013). In the case of bilingual language development, positive transfer may counteract the effect of limited input. Positive transfer refers to a statistical relationship between proficiencies in the two languages of a bilingual child that may reflect that a child can apply proficiencies developed in one language also to the other language (Wang, Perfetti, & Liu, 2005; Wang, Park, & Lee, 2006; Verhoeven, 2007; Scheele, Leseman, & Mayo, 2010). It is facilitated by linguistic overlap between languages, but children may also transfer common underlying proficiencies regardless of linguistic overlap; examples of common underlying proficiencies are phonological awareness, conceptualization, metacognitive and metalinguistic strategies, and pragmatic aspects of language use ('Interdependence hypothesis' Cummins, 1979, 1991, 2000).

Input differs between bilinguals and monolinguals, but variation in input also exists between bilinguals. Investigating variation within a group of bilinguals is important because bilinguals are highly heterogeneous (Dixon, Wu, & Daraghmeh, 2012; Grosjean & Li, 2012; Luk & Bialystok, 2013) and is relevant for understanding why some bilingual children are more at risk for language and literacy delays than others (Hammer, Hoff, Uchikoshi, Gillanders, Castro, & Sandilos, 2014). Within a sample of bilinguals, one factor that may impact on language development is SES, as a higher SES has been found to be associated with more language input (quantity) and with a higher quality of language input (Hart & Risley, 1995; Huttenlocher, Vasilyeva, Waterfall, Vevea, & Hedges, 2007; Golberg, Paradis, & Crago, 2008; Rowe & Goldin-Meadow, 2009; Scheele et al., 2010; Prevoo, Malda, Mesman, Emmen, Yeniad, van IJzendoorn, & Linting, 2013), although there is considerable variation in input within low SES families (Weisleder & Fernald, 2013), and the same will be true for mid and high SES families. Regarding SES, it is relevant to note that bilingualism is often not entirely independent of SES, as country-specific migration policies may result in high numbers of migrants with relatively high SES or relatively low SES. In addition to SES, language status may have an impact on input. Language status determines intergeneration transmission of home languages and the availability of resources such as children's books, newspapers, and Internet, which are limited for home languages with low status and no history of literacy (Scheele et al., 2010); low status can thus lead to reduced input quantity.

Below, we will turn to input, and factors that co-determine the (effect of) input to bilingual children such as transfer, SES and language status, in relation to bilingual children's lexical and narrative development.

1.2 Comparisons of bilingual and monolingual children: Lexicon

Several studies have shown that both input quantity and quality are associated with bilingual children's lexical outcomes (Pearson, Fernández, Ledeweg, & Oller, 1997; Hoff et al., 2012; Hoff & Core, 2013). Lower levels of input predict that bilingual children have weaker lexical skills in one language than monolingual children learning the same language. This hypothesis is confirmed in the large-scale study of Bialystok and colleagues for lexical comprehension (2010), but other researchers observed no differences for lexical comprehension and production between monolingual and bilingual children (De Houwer et al., 2013; Pearson, Fernández, & Oller, 1995). This suggests that the language input of bilingual children varies greatly and that there is a group of bilingual children whose input in one language is sufficient to reach a monolingual level (De Houwer, 2007).

A few studies have demonstrated relationships between SES, home language use and children's vocabulary knowledge (Golberg et al., 2008; Prevoo et al. 2013). For example, in their study with 6-year old Turkish-Dutch children from lower SES backgrounds, Prevoo and colleagues (2013) observed a negative relationship between SES and use of Turkish, showing that mothers with a lower SES used more Turkish than mothers with a higher SES. In their study, SES was positively correlated with reading input in Dutch. Consequently, in this overall low SES sample, the children from higher SES mothers knew fewer words in Turkish and more words in Dutch compared to children raised by lower SES mothers.

Observing different patterns across two bilingual migrant communities, Scheele et al. (2010) pointed to the role of orthography and language status as factors that mediate the relationship between input and children's lexical comprehension. They compared bilingual Turkish-Dutch and Berber/Tarifit-Dutch 3-year-old preschoolers and found that both bilingual groups scored lower than Dutch monolinguals on Dutch lexical comprehension. However, the largest gap was found for the Turkish-Dutch group. In contrast to Berber languages, Turkish is a written language and has high status within the Turkish community. Both orthography and status may support home language use in the Turkish families at the expense of Dutch input, influencing children's lexical development in Dutch.

Effects of input and home environment may be reduced if children can transfer knowledge and experience. Uccelli and Páez (2007) did not find significant relationships between the two (expressive) vocabulary scores of Spanish-English 4-year-old children, however. A recent study on Turkish-Dutch 5- and 6-year-old children demonstrated that transfer effects are found for children who used more Turkish than Dutch, and not for children in the reverse circumstances (Prevoo, Malda, Emmen, Yemiad, & Mesman, 2015). These results suggest that transfer of lexical knowledge happens but is limited, probably because word forms tend to vary greatly across languages even though concepts are shared. The impact of cross-linguistic influence on lexical transfer is confirmed by research showing that cognates facilitate the lexical development of bilingual children (Schelletter, 2002).

1.3 Comparisons of bilingual and monolingual children: Narration

Narrative skills allow children to tell and understand stories, and can be analyzed in different ways. Microstructural analyses focus on lexical and grammatical elements, especially those elements that establish links across sentences, such as referential expressions or connectives. Macrostructural analyses "examine children's language skills beyond the utterance level and document children's ability to relate concepts that transcend the individual utterance" (Heilmann, Miller, & Nockerts, 2010: 154). Macrostructure refers to the higher-order hierarchical organization of narratives; this includes a story's episodic structure (Heilmann et al., 2010), which can be captured in sequences consisting of goal (what a character aims to accomplish), attempt (a character's efforts to achieve the goal) and outcome (whether the character reached the goal) (Westby, 2005). Internal state language, another component of narration that exceeds the utterance level, indicates a child's understanding and awareness of intentionality and goal-directed behavior of characters (Nippold, Ward-Lonergan, & Fanning, 2005). While microstructure tends to be language-specific, macrostructure is considered to be universal (Berman & Slobin, 1994; Simon-Cereijido & Gutiérrez-Clellen, 2009; Iluz-Cohen & Walters, 2012; Gagarina, Klop, Tsimpli, & Walters, 2016).

While previous research shows that input and home environment influence bilingual children's lexicon, this could be different for narrative macrostructure (Pearson, 2002; Fiestas & Peña, 2004; Iluz-Cohen & Walter, 2012; Hipfner-Boucher et al., 2015). As narrative macrostructure is considered to be more universal, children's outcomes on measures of narrative macrostructure could be relatively independent of input in a specific language. Hipfner-Boucher and colleagues (2015) compared three groups of children who all learned English: monolingual children, bilingual children who were mostly exposed to English at home, and bilingual children mostly exposed to a minority language at home. Analyses of children's retelling of stories showed that the bilinguals with a large amount of English exposure patterned similarly to the monolinguals on microstructure, and outperformed bilinguals with a large amount of minority language exposure. No between-group differences emerged for narrative macrostructure, confirming that narrative macrostructure production is relatively immune to input effects.

The study by Rodina (2017) shows that between-group differences may be modulated by how narrative macrostructure is measured. In this study no differences emerged between bilingual and monolingual children on narrative macrostructure for both the majority (Norwegian) and minority (Russian) language of bilingual children based on comprehension questions. Narrative macrostructure *production* of bilingual children did show lower performance in comparison to monolinguals, however, but only in the minority language, which was the weaker language in which the children received less input. In this study, production was not measured by retelling as was done in the study by Hipfner-Boucher et al. (2015), but through story generation. Narrative comprehension is also modulated by task properties. For example, using a similar narrative instrument as Rodina (2017), Maviş, Tunçer, and Gagarina (2016) found that children answered questions about the story by the experimenter more accurately than questions asked about a story they told themselves (see Otwinowska, Mieszkowska, Białecka-Pikul, Opacki, & Haman (2018) for similar results).

In accordance with findings showing that bilingual children perform well on narrative macrostructure comprehension (Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Hipfner-Boucher et al., 2015; Rodina, 2017) and in line with Cummins' Interdependence hypothesis (Cummins, 1979, 1991, 2000) are various observations which suggest that bilingual children transfer narrative macrostructure proficiencies across languages. Spanish-English bilingual children show cross-language correlations for narrative macrostructure (Pearson, 2002; Uccelli & Páez, 2007). In addition, Uccelli and Páez (2007) observed that Spanish story structure in kindergarten predicted first-grade English narrative quality even when the effects of English vocabulary and English narrative productivity were controlled.

Migrant children can only benefit from transfer if they have well developed narrative proficiency in the home language. From this point of view, SES, which is linked to quantitative and qualitative aspects of parental input (Hoff, 2006) in any language, could influence children's possibilities to transfer home language proficiencies to the majority language. It has been found that children raised in low SES families are outperformed by children from high SES backgrounds on narrative macrostructure, although the effect of SES on narrative microstructure was larger (Pearson, 2002). Narrative macrostructure may thus be a partially learned skill that requires examples and input, although independent of a specific language. Consequently, low SES children may have less to transfer than high SES children.

1.4 The current research

The current research resembles the work by Uccelli and Páez (2007) who also investigated bilingual children's lexical and narrative development over time. Whereas Uccelli and Páez' study was focused on lexical and narrative production in Spanish-English bilingual children, the current research is about the lexical and narrative comprehension of bilingual Tarifit-Dutch and Turkish-Dutch children who are raised in families that migrated to the Netherlands from Morocco and Turkey. Tarifit is a Berber or Rifian language, spoken in the north of Morocco. Previous research demonstrated that findings for narrative macrostructure production do not necessarily generalize to comprehension (Rodina, 2017), and that narrative comprehension outcomes are dependent on type of task (Maviş et al., 2016). To investigate lexical comprehension, we administered the widely used Peabody Picture Vocabulary Test (Dunn & Dunn, 1997; Schlichting, 2005). Children's narrative macrostructure comprehension was investigated with the Multilingual Assessment Instrument for Narratives (MAIN) (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012; Gagarina, Klop, Tsimpli, & Walters, 2016) by analyzing children's responses to comprehension questions about a story told by the experimenter (listening comprehension), and children's responses to comprehension questions about a second story they told themselves (generated story comprehension).

This study was guided by three research questions. First, we wanted to know: *Does Dutch lexical and narrative comprehension differ between bilingual and monolingual children?* As argued above, narrative macrostructure comprehension requires less language-specific knowledge than lexical comprehension because narrative macrostructure is, at least in part, considered universal and hence independent of a specific language. Therefore, we predicted that the bilingual children would be outperformed by the monolingual children on Dutch lexical comprehension but that the groups would perform similarly on Dutch narrative comprehension. We expected the same between-group pattern for listening comprehension and generated story comprehension, but overall higher scores on listening comprehension, in line with Maviş et al. (2016). Migrant children are typically dominant in the minority language at preschool ages and become dominant in the majority language during elementary school (Extra et al., 2001). The bilingual children in this study went to elementary school already at time 1, hence their exposure to Dutch was frequent and accumulated during the three times that the children were tested. Consequently, it is expected that differences between bilinguals and monolinguals would be smaller at time 3 than time 1 (and time 2 is expected to be in between).

The second research question was formulated to explore differences between two bilingual migrant communities: *To what extent do Tarifit-Dutch and Turkish-Dutch children differ in their Dutch lexical and narrative comprehension?* Research by Scheele et al. (2010) with 3-year-old toddlers demonstrated that in Tarifit-speaking families relatively more Dutch and less home language was used compared to Turkish families, and these differences in input and home environment were reflected in the children's Dutch receptive vocabularies. In the Netherlands, children start full day kindergarten at the age of 4 years. The children in the present study had attended kindergarten for at least 12 months at the first wave of data collection. We expected differences between Tarifit-Dutch and Turkish-Dutch children in Dutch lexical comprehension to be less prominent than in the study by Scheele et al. (2010), because the children in our study had spent a significant amount of time in a school environment where only Dutch is used, reducing between-group differences related to differences in the use of Dutch at home. In addition, we expected any differences to diminish over time. We did not expect any differences between the groups due to typological factors as Tarifit and Turkish are both very different from Dutch and have a similar distance from Dutch. This can be illustrated through a comparison of the average normalized Levenshtein distances. The Levenshtein distance is an index of the number of manipulations (changes and additions) needed to get from the shorter to the longer word form in two languages: the fewer manipulations are needed the closer two languages are. Normalized Levenshtein distances are controlled for length of the strings. The normalized Levenshtein distance is 99 for the Tarifit-Dutch language pair and 101 for Turkish-Dutch (Automated Similarity Judgment Program, asjp62; Bakker et al., 2009; Wichman, Holman, & Brown, 2016). There were no specific between-group differences expected for narrative comprehension, for reasons explained above.

The third research question pertained to the impact of input and home environment on bilingual children's lexical and narrative abilities in Dutch: *Are input factors at home (language use, language richness, parental SES) related to Tarifit-Dutch and Turkish-Dutch children's lexical and narrative comprehension in Dutch?* Available research suggests that children's narrative macrostructure is less amenable to language-specific input effects than their lexicon. We expected that Dutch language use and language richness would be positively related to lexical outcomes in Dutch. Language use and language richness in the home language (Tarifit, Turkish) were expected to be negatively correlated with children's Dutch lexical outcomes, as more home language use will imply less use of Dutch (Scheele et al., 2010). We furthermore expected that SES would be positively related to both children's lexical (Prevoo et al., 2012) and narrative macrostructure (Pearson, 2002) development in Dutch, but that relationships would be stronger for lexical development (Pearson, 2002). We formulated general predictions, but it cannot be excluded that different patterns emerge for the two bilingual subgroups. As indicated above, previous research indicated that input patterns in the homes of Tarifit-Dutch children and Turkish-Dutch children are different (Scheele et al., 2010). Consequently, relationships with lexical and narrative comprehension found for one bilingual group may not generalize to the other.

2. Method

2.1 Participants

For the present study, data from 114 Tarifit-Dutch, Turkish-Dutch and monolingual Dutch children were analyzed most of whom were 5 or 6 years old at the first testing time. Assignment to either the bi- or monolingual group was based on parental report using the *Questionnaire for Parents of Bilingual Children* (PaBiQ) (Tuller, 2015). A child was considered monolingual if both parents always spoke Dutch to the child. A child was assigned to the bilingual group if at least one parent was a native speaker of the home language (Tarifit, Turkish), and spoke their native tongue with the child for an extensive period of the child's life (more detailed information on language use at home can be found in the Results' section, Table 5). The bilingual children were born in the Netherlands, and second or third generation migrant. From the age of 4 years they went to kindergarten, which implies that from the age of 4 years all children spent on average 24 hours a week during week days in a Dutch language environment. All parents reported that their child had received a considerable amount of exposure to Dutch before age 4 at home and/or in preschool programs and nurseries: on average 44% of this time (SD = 9), ranging between 29%

and 67%. For some children, their parents indicated that they received exposure to a third language. This was mostly Arabic, and was so infrequent (< 2% of the time) that the children cannot be considered trilingual. The bilingual group included 38 Tarifit-Dutch children and 31 Turkish-Dutch children, and the monolingual group included 45 children. One bilingual child was diagnosed with a language disorder at time 3, and for this reason removed from the study. None of the children included in this study had a clinical diagnosis.

Data were collected at three times with one year between each testing time. The drop-out rate was minimal (< 3% of the total sample). For 34 Tarifit-Dutch, 21 Turkish-Dutch and 43 monolingual Dutch children complete datasets were available that included measures of lexical and narrative comprehension at all three times. At the second testing time, additional Turkish-Dutch children were recruited to have more equal numbers of Tarifit-Dutch and Turkish-Dutch children. For these additional Turkish-Dutch children, data from only two testing times was therefore available. To include the data from all 114 children, we conducted separate analyses for the three times and refrained from performing longitudinal analyses; the data-analytic strategy is explained below.

Table 1 provides information on the children's ages in months, nonverbal IQ measured with the short version of the *Wechsler Nonverbal Scale of Ability* (Wechsler & Naglieri, 2008) and SES measured by PaBiQ and based on the average educational level of the mother and the father measured on a 9-point scale (see below for further details). In the monolingual sample, the majority of the parents (64%) scored 7–9 on the 9-point scale (referring to higher secondary education, higher professional education, or university degree). In the bilingual sample, the majority of the parents (63%) scored below 7 on the 9-point scale (referring to (pre-)vocational education, elementary education, or no formal education).

Table 1. Number of children, mean age in months (standard deviation), mean nonverbal intelligence (standard deviation) and mean socio-economic status (standard deviation) in monolingual and bilingual groups

	N	Age time 1	Age time 2	Age time 3	NVIQ	SES
Bilingual	69	67 (7)	79 (7)	91 (7)	98 (14)	5 (2)
Tar-Du	38	65 (7)	77 (7)	89 (7)	94 (14)	5 (2)
Tur-Du	31	69 (7)	81 (6)	94 (6)	102 (13)	5 (2)
Monolingual	45	71 (8)	82 (8)	94 (8)	107 (15)	7 (2)

Note. NVIQ = nonverbal intelligence measured with the Wechsler Nonverbal Scale of Ability, standard score (M = 100, SD = 15); SES = socio-economic status indexed by the average educational level of both parents, average of both parents measured on a 9-point scale with 1 = no formal education and 9 = university degree (not available for 12 bilingual children).

The monolingual sample contained 20 girls and 25 boys and the bilingual sample contained 35 girls and 34 boys. A 2x2 chi-square test showed that the distributions of boys and girls did not differ across the two groups ($\chi^2(1) = .43$, $p = .51$). The monolingual children were slightly older than the bilinguals at time 1, $F(1, 110) = 6.66$, $p = .01$, $\eta_p^{\,2} = .06$, time 2, $F(1, 110) = 5.91$, $p = .02$, $\eta_p^{\,2} = .05$, and time 3, $F(1, 103) = 5.53$, $p = .02$, $\eta_p^{\,2} = .05$. Nonverbal IQ scores were higher for the monolinguals compared to the bilinguals, $F(1, 112) = 10.48$, $p = .002$, $\eta_p^{\,2} = .09$, as were outcomes for SES as indicated by an independent samples Mann-Whitney U test ($U = 727$, $p < .001$).

The Tarifit-Dutch sample contained 20 girls and 18 boys and the Turkish-Dutch sample contained 15 girls and 16 boys; the distributions did not differ across the two bilingual groups ($\chi^2(1) = .12$, $p = .73$). The Turkish-Dutch children were slightly older than the Tarifit-Dutch children at time 1, $F(1, 65) = 5.30$, $p = .03$, $\eta_p^{\,2} = .08$, time 2, $F(1, 65) = 6.45$, $p = .02$, $\eta_p^{\,2} = .09$, and time 3, $F(1, 59) = 10.36$, $p = .002$, $\eta_p^{\,2} = .15$. Nonverbal IQ scores were higher for the Turkish-Dutch than the Tarifit-Dutch children, $F(1, 67) = 6.34$, $p = .014$, $\eta_p^{\,2} = .09$. The two bilingual groups did not differ in SES ($U = 727$, $p = .66$).

2.2 Measures and materials

2.2.1 *Lexical comprehension*

Receptive vocabulary in Dutch was measured with the Peabody Picture Vocabulary Task (PPVT-III-NL; Schlichting, 2005). The PPVT is a standardized receptive vocabulary test designed for the age range from 2 years and 3 months to 90 years. It contains 204 items divided over 17 sets. The sets are ordered according to difficulty and each set consists of twelve items. In this task, a child hears a stimulus word and has to choose the correct referent out of four pictures. The PPVT was administered and scored according to the official guidelines. This means that the starting set was determined by a child's age and the task was terminated after a child produced nine or more errors within one set. For the analyses, raw PPVT scores were used.

2.2.2 *Narrative comprehension*

In this study, the *Multilingual Assessment Instrument for Narratives* (MAIN; Gagarina et al., 2012) was used. This narrative task was created within the framework of the COST Action IS0804 (Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment). MAIN is an assessment tool that is suitable for children from 3 to 10 years old and can be used for the evaluation of both comprehension and production of narratives in both languages of a bilingual child. MAIN contains four stories (Cat, Dog, Baby birds, Baby goats), each with a carefully designed, full-colour, six-picture sequence. The four stories are controlled for cognitive and linguistic complexity, parallelism in macrostructure and microstructure, as well as cultural appropriateness and robustness (Gagarina et al., 2012, p. 1). The parallel stories consist of three episodes and are of equal length. Each episode introduces one or more characters and allows for the description of internal states (e.g., curious, playful) of the character(s). Furthermore, each episode contains a goal (e.g, Cat wanted to catch the Butterfly), an attempt (e.g, Cat jumped forwards) and an outcome (e.g., Cat fell into the bush or Butterfly escaped). The current study used the Dutch MAIN version.

The narrative elicitation procedure can influence comparisons between monolinguals and bilinguals (Gutiérrez-Clellen, 2002). The design of MAIN with four stories allows the experimenter to use model story, retelling or story generation as the elicitation procedure. In the present study, a combination of model story and story generation was employed. The model story provides contextual support for the child to tell a story and enables testing listening comprehension. At the same time, the story generation task is less dependent on memory than a recall task. The children first heard the model story, which was followed by ten comprehension questions that targeted the goal of each episode (3 points), the internal states of the episode character (6 points), and a final question that also pertained to an internal state but that furthermore required a child to make an inference based on the story as a whole (1 point) (see Appendix). Subsequently, they were shown the six pictures of a new story and were asked to generate a story. After this, another ten comprehension questions were asked, which were structured in the same way as the questions about the model story.

Half of the children in the monolingual sample and half of the children in the bilingual sample had in Dutch a combination of Cat (listening comprehension) and Baby birds (story generation). The other half had a combination of Dog (listening comprehension) and Baby goats (story generation). Those bilingual children who had Cat (listening comprehension) and Baby birds (story generation) in Dutch, had Dog (listening comprehension) and Baby goats (story generation) in the home language, and vice versa. For the present study, we did not consider the home language narratives. The order was fixed for all bilingual children: the home language was tested first and Dutch was tested one week later to eliminate the chance of learning effects from Dutch – the expected dominant language of most children – to the home language.

In this study, we focused on the comprehension questions. The reason is that our main aim was to compare language development across domains, and only in the receptive modality both lexical and narrative measures were available. All narratives were recorded with a highly sensitive microphone (Samson Go Mic) and were scored offline. A previous study with partly the same sample (a subset of the monolinguals and bilinguals) revealed high agreement between two independent coders with inter-class coefficients of respectively .93 and .89 for the listening comprehension and comprehension after story generation (Boerma et al., 2016).

2.2.3 *Home language environment*

The PaBiQ (Tuller, 2015) provided information on language use and language richness in the children's home environments and SES. The PaBiQ is a short version of a longer questionnaire piloted by research groups in several countries within COST Action IS0804, which was in part based on the ALEQ (Paradis, 2011) and the ALDeQ (Paradis et al., 2010). Dutch language use indicates the percentage of time a child was addressed in Dutch and was calculated as the mean percentage of Dutch input that the child received from his mother, father, siblings and other adults who took care of the child at least once per week. For each of these people we asked how often (s)he spoke Dutch to the child: 'never' (0%), 'seldom' (25%), 'sometimes' (50%), 'usually' (75%) or 'always' (100%), and then we averaged the answers. The same was measured for home language use. This resulted in one value for Dutch language use and one value for home language use. Dutch language richness is a more qualitative measure that looks at how often Dutch is used with family friends and peers, as well as during reading activities, watching television/movies, and oral storytelling, each measured on a three-point scale ranging from 0 = very infrequently to 2 = every day. The same was measured for home language richness. For the variable language richness, all points for language use outside the home and different language activities were added up and divided by the maximum score that was possible (8). This resulted again in one richness value per language.

SES was calculated as the average educational level of the mother and the father, as measured on a 9-point scale: (1) no education, (2) elementary education, (3) pre-vocational secondary education, (4) pre-vocational secondary education, theoretical programme, (5) vocational education, level 2 or 3, (6) vocational education, level 4, (7) higher secondary education, (8) higher professional education, (9) university degree. Parental education is the most commonly used indicator of SES in research with children (Ensminger & Fothergill, 2003) and is predictive of other SES indicators such as income and profession (Bornstein, Hahn, Suwalsky, & Haynes, 2003).

2.3 Procedures

This research was screened by the Standing Ethical Assessment Committee of the Faculty of Social and Behavioral Sciences at Utrecht University. Criteria were met and further verification was not deemed necessary. Parents of participants gave informed consent. All participants were tested individually in a quiet room at their school. They completed a battery of tests that included, in addition to vocabulary and narratives, several other tasks testing language, working memory and attention. The monolingual children were tested in two separate sessions, each lasting approximately one hour. The PPVT-III-NL and MAIN in Dutch were completed in the second session as the third and sixth tasks respectively. The bilingual children were tested in three sessions, of which the first was a session in the home language and the second and third were in Dutch. Just like the monolingual children, the bilingual children completed the PPVT-III-NL and MAIN in Dutch as the third and sixth tasks in the final session. Both tasks in the home language were completed in the first session. The first and the third test sessions were at least one week apart to reduce the chance of learning effects from the home language to Dutch. The parental questionnaire was administered during a telephone interview with one of the child's parents. The interview was conducted by bilingual assistants who were proficient in both Dutch and Tarifit/Turkish and could therefore be carried out in the preferred language of the parent.

2.4 Data-analysis

2.4.1 *Research question 1*

We compared bilingual and monolingual children on lexical and narrative development using analyses of variance with Group (bilingual, monolingual) as the independent variable and PPVT (raw scores), listening comprehension (MAIN1, proportion correct), and generated story comprehension (MAIN2, proportion correct) as dependent variables. In a few cases, research assistants, erroneously, did not ask all ten questions. The percentage of questions that were not asked out of all questions that should be asked was less than 1%. Omission of questions concerned 8 monolingual (2.99%) and 20 bilingual children (5.21%) with no bias towards one of the two groups ($\chi^2(1) = 1.90$, $p = .19$). To correct for the omitted questions and avoid depressing the children's scores because of experimenter errors, we calculated proportions correct with the total number of questions asked as the denominator. These adjusted proportions are only minimally different from proportional scores based on a denominator of 10 (= total number of questions that should be asked), as there were relatively few omitted questions. Hence the proportional scores we report closely reflect the number of correct items (i.e. a proportion of .86 reflects that on average about 8.6 questions received a correct answer).

To be able to include a maximum amount of data points, we ran separate analyses for each time instead of including time as a within-subject factor using a repeated measures analysis of variance. This implies that the analyses do not account for the correlated nature of the data. The data set was not sufficiently large for growth curve analyses. In order to evaluate any changes over time, differences in effect size between time 1, time 2 and time 3 were interpreted. For this, we followed the rule of thumb that a $\eta_p^2 = .01$ refers to a small effect, $\eta_p^2 = .06$ refers to a medium effect, and $\eta_p^2 = .14$ refers to a large effect (Richardson, 2011). In cases where normality could be assumed, ANCOVAs were performed with age and NVIQ as a covariate because the monolinguals were slightly older and had a higher NVIQ scores than the bilinguals, as described earlier. The monolinguals also had higher SES. SES was not available for 12 of the bilingual children. For this reason, we did not include SES in analyses as a covariate *unless* adding SES as a covariate led to different results. In cases where normality could not be assumed, a Mann Whitney U test was performed. Because this test does not allow for the inclusion of covariates, any statistically significant between-group differences were checked based on a subsample matched on age and NVIQ.

In addition, we explored whether there were any differences between bilingual and monolingual children based on the types of questions. To this end, the questions were grouped into three categories: goals, internal states and inferences. This was done for both listening comprehension and generated story comprehension at all three times. To determine whether there were any differences between the two groups in how well they did on each question category, χ^2 tests were applied.

2.4.2 *Research question 2*

Between-group comparisons were performed to compare the two bilingual groups (Tarifit-Dutch, Turkish-Dutch), using the same strategy as for research question 1.

2.4.3 *Research question 3*

Correlational analyses (Spearman rank correlations) were performed within the two bilingual groups between the input measures Dutch language use, Dutch language richness, Home language use, Home language richness, SES, on the one hand, and lexical (PPVT) and narrative comprehension (MAIN1, MAIN2) scores on the other hand. Following Cohen's (1988) guidelines, correlations with a strength of .50 were considered strong, .30 were considered moderate and .10 were considered weak. Both Dutch and Home language use and richness were investigated; these values were not entirely complementary, because some children received (limited) exposure to a third language (see Participants' section).

The analyses to answer the three research questions comprised multiple comparisons between the same groups and assessing multiple relations within the same groups, which may lead to an increase of a type 1 error. Because separate comparisons for lexical and narrative measures were planned, no Bonferroni corrections are needed to correct for multiple comparisons. We did apply a Bonferroni correction to correct for the increase of comparisons because analyses were conducted for time 1, time 2 and time 3 separately (α decision level = .05/3 = .015), in order to prevent type 1 errors but also avoid being unnecessarily stringent and increasing the risk of a type 2 error (Armstrong, 2014).

3. Results

3.1 Research question 1: Comparing bilingual and monolingual children

The results of the lexical and narrative measures at time 1, time 2 and time 3 are summarized in Table 2 for the bilingual and monolingual group.

Table 2. Lexical and narrative development in the bilingual and monolingual samples, means (standard deviations)

	Bilingual			Monolingual		
	PPVT	MAIN1	MAIN2	PPVT	MAIN1	MAIN2
Time 1	73 (11)	.86 (.15)	.69 (.19)	88 (12)	.92 (.09)	.83 (.16)
Time 2	84 (12)	.91 (.11)	.80 (.17)	100 (12)	.93 (.08)	.86 (.14)
Time 3	93 (10)	.94 (.08)	.90 (.13)	106 (11)	.91 (.09)	.91 (.15)

Note. PPVT = Peabody Picture Vocabulary Test (lexicon comprehension), raw scores; MAIN1 = Multilingual Assessment Instrument for Narratives (listening comprehension), proportion correctly answered questions; MAIN2 = Multilingual Assessment Instrument for Narratives (generated story comprehension), proportion correctly answered questions; for most children, the percentage is based on *n* out of 10 questions, hence it can be assumed that an average proportion of .86 refers to an average of about 8.6 items correct.

3.1.1 *Lexical comprehension*

An ANCOVA with Group (monolingual, bilingual) as the between-subjects variable and Age and NVIQ as covariate and raw PPVT scores as the dependent variable showed that the monolingual children performed better than the bilingual children at time 1, $F(1, 102) = 25.39$, $p < .001$, $\eta_p^2 = .20$, time 2, $F(1, 108) = 31.81$, $p < .001$, $\eta_p^2 = .23$, and time 3, $F(1, 101) = 19.46$, $p < .001$, $\eta_p^2 = .16$. Age was a significant covariate at time 1, $F(1, 102) = 35.64$, $p < .001$, $\eta_p^2 = .26$, time 2, $F(1, 108) = 37.68$, $p < .001$, $\eta_p^2 = .26$, and time 3, $F(1, 101) = 15.26$, $p < .001$, $\eta_p^2 = .13$. NVIQ was a significant covariate at time 1, $F(1, 102) = 9.21$, $p = .003$, $\eta_p^2 = .08$, time 2, $F(1, 108) = 12.65$, $p < .001$, $\eta_p^2 = .11$, and time 3, $F(1, 101) = 20.69$, $p < .001$, $\eta_p^2 = .17$. Analyses that also included SES as a covariate, in addition to Age and NVIQ, produced similar results with respect to the between-group difference. These analyses showed a decreasing effect size (time 1 $\eta_p^2 = .16$; time 2

$\eta_p^2 = .17$; time 3 $\eta_p^2 = .10$). The between-group differences are significant with and without a Bonferroni correction for multiple comparisons. To further verify the between-group differences, a matched-group analysis was performed which included subgroups of 40 monolingual and 40 bilingual children for whom full datasets were available and that were matched on NVIQ, $F(1, 78) = 1.31$, $p = .26$, and Age, $F(1, 78) = .70$, $p = .41$. The monolinguals outperformed the bilinguals on lexicon comprehension at time 1, $F(1, 78) = 23.23$, $p < .001$, $\eta_p^2 = .23$, time 2, $F(1, 78) = 78.74$, $p < .001$, $\eta_p^2 = .27$, and time 3, $F(1, 78) = 18.16$, $p < .001$, $\eta_p^2 = .19$.

3.1.2 *Narrative comprehension*

The children in both groups performed accurately on the narrative comprehension questions. Due to the children's high accuracies, MAIN scores deviated from normality and for this reason non-parametric tests were employed. We first investigated whether performance on MAIN1 (listening comprehension) and MAIN2 (generated story comprehension) was different (Maviş et al., 2016). A Wilcoxon Signed Ranked test revealed better performance on MAIN1 than MAIN2 at time 1 ($W = 460$, $p < .001$) and time 2 ($W = 700$, $p < .001$). At time 3, the difference did not reach significance ($W = 966$, $p = .095$). These results indicate that separate treatment of the two outcomes is warranted. Mann-Whitney U tests with Bonferroni correction revealed no differences between bilinguals and monolinguals on MAIN1 at time 1, time 2 and time 3. On MAIN2, the monolinguals outperformed the bilinguals at time 1 ($U = 843$, $p < .001$). The between-group difference at time 1 could be related to the lower NVIQ and younger age of the bilingual group, which both correlate significantly with MAIN2 scores at time 1 (NVIQ: $r_s(110) = .27$, $p = .004$; Age: $r_s(110) = .39$, $p < .001$). An analysis with the subgroups of 40 bilingual children and 40 monolinguals matched on NVIQ and Age showed that the associated p-value is higher than the Bonferroni corrected α decision level (.015), suggesting no difference ($U = 688$, $p = .017$). At times 2 and 3 the two groups did not differ significantly from each other.

MAIN comprehension questions asked about goals, internal states and required children to make inferences. Table 3 shows how bilingual and monolingual children responded to the three different question types. The proportions relative to the total number of points for all question types within a group at a certain time are shown. Points per question are listed in the Appendix. The proportions demonstrate very similar distributions for the monolinguals and bilinguals. None of the six χ^2 tests (computed based on absolute points) returned a significant result, neither at the α decision level of .05 nor at the more stringent .015 level, confirming that the groups overall showed few differences on narrative comprehension.

Table 3. Proportions of points per question type (goals, internal states, inferences) in the bilingual and monolingual samples for listening comprehension/generated story comprehension

	Bilingual			Monolingual		
	Goals	Internal states	Inferences	Goals	Internal states	Inferences
Time 1	.33/.35	.60/.54	.07/.11	.32/.33	.62/.56	.06/.10
Time 2	.31/.33	.60/.57	.09/.10	.32/.34	.62/.54	.06/.11
Time 3	.32/.33	.61/.57	.07/.11	.33/.32	.63/.58	.05/.10

3.2 Research question 2: Comparing bilingual Tarifit-Dutch and Turkish-Dutch children

The results on the lexical and narrative measures at time 1, time 2 and time 3 are summarized in Table 4 for the two bilingual groups separately.

Table 4. Lexical and narrative development in the Tarifit-Dutch and Turkish-Dutch samples, means (standard deviations)

	Bilingual Tar-Du			Bilingual Tur-Du		
	PPVT	MAIN1	MAIN2	PPVT	MAIN1	MAIN2
Time 1	73 (11)	.83 (.15)	.66 (.19)	73 (12)	.89 (.14)	.74 (.19)
Time 2	83 (12)	.93 (.10)	.74 (.19)	84 (12)	.90 (.12)	.88 (.12)
Time 3	93 (9)	.94 (.08)	.87 (.12)	94 (13)	.94 (.09)	.93 (.13)

Note. PPVT = Peabody Picture Vocabulary Test (lexicon comprehension), raw scores; MAIN1 = Multilingual Assessment Instrument for Narratives (listening comprehension), proportion correctly answered questions; MAIN2 = Multilingual Assessment Instrument for Narratives (generated story comprehension), proportion correctly answered questions; for most children, the percentage is based on *n* out of 10 questions, hence it can be assumed that an average proportion of .86 refers to an average of about 8.6 items correct.

3.2.1 *Lexical comprehension*

An ANCOVA with Group (Tarifit-Dutch, Turkish-Dutch) as the between-subjects variable, Age and NVIQ as covariates and raw PPVT scores as the dependent variable did not return a significant effect of Group at time 1, time 2, or time 3, indicating that the Tarifit-Dutch and Turkish-Dutch children children did not differ in comprehending Dutch words.

3.2.2 *Narrative comprehension*

Nonparametric Mann Whitney U tests showed no differences between the two groups for MAIN1 scores at time 1, time 2 and time 3, showing that children in the two groups did not differ in listening comprehension. The descriptive statistics in Table 6 suggest that the Turkish-Dutch children had somewhat higher MAIN2 outcomes. This difference, however, only reached significance at time 2 ($U = 288$, $p = .002$). The Tarifit-Dutch children were younger and scored lower on NVIQ than the Turkish-Dutch children, and these differences may explain Tarifit-Dutch children's lower accuracy. In the bilingual sample, Age at time 2 was not significantly correlated with MAIN2 scores ($r_s(65) = .23$, $p = .072$), but the correlation with NVIQ did reach significance ($r_s(65) = .25$, $p = .047$). An analysis which included a subgroup of 25 bilingual Tarifit-Dutch children and 25 bilingual Turkish-Dutch children who did, on average, not differ in NVIQ ($F(1, 48) = 1.13$, $p = .720$) and Age ($F(1, 48) = 3.42$, $p = .071$) however confirmed the lower performance of the Tarifit-Dutch children on MAIN2 at time 2 ($U = 181$, $p = .009$). No significant between-group differences emerged for MAIN2 at times 1 and 3. Recall that for all comparisons a Bonferroni-corrected α decision level of .015 was used.

3.3 Research question 3: The role of input in the bilingual samples

Table 5 provides the descriptive data about language use in the homes of the bilingual Tarifit-Dutch and Turkish-Dutch children.

Table 5. Language use at home, means (standard deviations) and 95% CI interval of the difference

	Bilingual Tar-Du	Bilingual Tur-Du	95% CI interval
D-Use	.59 (.10) [a]	.44 (.10)	[−.21−−.09]
D-Rich	.83 (.10)	.70 (.16)	[−20−−.06]
H-Use	.37 (.09) [a]	.56 (.13)	[.13–.24]
H-Rich	.19 (.11)	.57 (.21)	[.30–.46]

Note. D-Use = use of Dutch at home; D-Rich = Dutch richness at home; H-Use = use of the home language (Tarifit, Turkish) at home; H-Rich = home language (Tarifit, Turkish) richness at home.
a. D-Use and H-Use do not add up to 1.00 because in a few cases, parents indicated use of a third language; this was only 2% of the time, hence the children cannot be considered trilingual.

For none of the four variables, zero was part of the 95% confidence interval of the difference indicating that Dutch use and richness scores were higher in the Tarifit-Dutch group than in the Turkish-Dutch group. Home language use and richness were higher in the Turkish-Dutch group compared to the Tarifit-Dutch group.

Table 6 provides an overview of the correlations between input variables, on the one hand, and lexical and narrative comprehension, on the other in the Tarifit-Dutch and Turkish-Dutch group, respectively.

Table 6. Correlations between lexical and narrative measures and input measures in the Tarifit-Dutch and Turkish-Dutch samples

Tarifit-Dutch	**Time 1**			**Time 2**			**Time 3**		
	PPVT	**MAIN1**	**MAIN2**	**PPVT**	**MAIN1**	**MAIN2**	**PPVT**	**MAIN1**	**MAIN2**
D-Use	.02	.08	.10	.06	−.42*	−.27	−.01	.01	.01
D-Rich	−.08	−.13	.10	−.09	−.40*	−.27	−.14	−.06	.08
H-Use	−.03	−.21	.07	−.13	.43*	.28	.05	.03	−.02
H-Rich	.06	**.44***	.17	.07	.32	.39*	.18	.26	.04
SES	−.02	.05	−.02	−.09	−.33	−.09	−.12	−.36*	−.19
Turkish-Dutch	**Time 1**			**Time 2**			**Time 3**		
	PPVT	**MAIN1**	**MAIN2**	**PPVT**	**MAIN1**	**MAIN2**	**PPVT**	**MAIN1**	**MAIN2**
D-Use	.44*	.35	.40	.46*	.10	.04	.44*	.20	−.01
D-Rich	.40	.29	.16	.38	−.03	.30	.30	.29	.09
H-Use	−.44*	−.35	−.40	−.46*	−.10	−.04	−.44*	−.20	.01
H-Rich	−.41	−.19	.03	−.32	−.19	−.03	−.48*	−.002	−.01
SES	.32	.18	.18	**.54***	.10	.35	**.60****	−.04	−.15

Note. D-Use = use of Dutch at home; D-Rich = Dutch richness at home; H-Use = use of the home language (Tarifit, Turkish) at home; H-rich = home language (Tarifit, Turkish) richness at home; SES = socioeconomic status indexed by the average educational level of both parents; PPVT = Peabody Picture Vocabulary Test (lexicon comprehension); MAIN1 = Multilingual Assessment Instrument for Narratives (listening comprehension); MAIN2 = Multilingual Assessment Instrument for Narratives (generated story comprehension)
* correlation is significant at the .05 level
** correlation is significant at the .01 level; correlations that are moderate (.30) or strong (.50) are marked through grey cells (following Cohen, 1988); correlations in bold are significant after Bonferroni correction.

The correlations in Table 6 show different patterns for the two bilingual groups. The only correlation in the Tarifit-Dutch group that reached significance (after Bonferroni correction) was the correlation between Home language richness and MAIN1 at time 1 ($p = .011$), showing that children with richer input at home in Tarifit scored higher on listening comprehension. In the Turkish-Dutch group, the correlations between Dutch use and Home language use, on the one hand, and PPVT at times 1, 2 and 3 were not significant anymore after Bonferroni correction. The correlation between SES and PPVT at times 2 and 3 remained significant ($p = .007$; $p = .004$).

To examine whether or not the relation between SES and Dutch lexical comprehension was mediated by Dutch use, partial correlations were performed between lexical comprehension at time 2 and time 3 and SES in which the effect of Dutch use was partialled out. Controlling for Dutch use weakened the strength of the relationship between SES and PPVT at both times. Both at time 2 and 3, the partial correlation was not significant anymore (time 2: $r(21) = .39$, $p = .069$; time 3: $r(18) = .42$, $p = .063$). The relation between SES and Turkish-Dutch children's comprehension of Dutch words is thus in part explained by Dutch use at home.

4. Discussion and conclusion

In this study we researched the majority language (Dutch) development of bilingual Tarifit-Dutch and Turkish-Dutch children who are raised in families that migrated from Morocco and Turkey to the Netherlands. Our main aim was to investigate the children's Dutch language comprehension across two domains: lexicon and narration. The bilingual group performed consistently less well than a monolingual control group on Dutch lexical comprehension, but the two groups hardly differed in their performance on Dutch narrative comprehension. Dutch lexical comprehension was unrelated to the input that the bilinguals received in their home environment, whereas home language use (in contrast with Dutch language use) and home language richness were positively associated with narrative comprehension. The bilingual Tarifit-Dutch and Turkish-Dutch subgroups performed the same on lexical comprehension. There were some indications that the Turkish-Dutch children were better at comprehending Dutch narratives than the Tarifit-Dutch group. Moreover, correlations between input at home and Dutch lexical and narrative comprehension were different across the two bilingual groups, although it must be noted that, out of the many correlations that were computed few were statistically significant. The results suggest, first, that bilingual language delays may not hold for all domains of language, and might be more prominent for the comprehension of (single) words than for children's understanding of (short) stories. Secondly, results found for one bilingual group do not necessarily generalize to another group of bilingual children.

4.1 Lexical and narrative comprehension: Bilinguals versus monolinguals

The lower lexical outcomes of the bilingual group parallel the often reported finding that bilingual children have smaller vocabularies (when measured in one language) compared to monolinguals (Bialystok et al., 2010), and reflects the distributed nature of bilingual children's lexical knowledge, which, in turn, reflects their distributed input (Oller & Pearson, 2002; Pearson et al., 1993). The lexical gap between bilinguals and monolinguals was found at all three times. Even though there are indications that the gap narrowed, the effect size at time 3 was still medium-sized (at best), suggesting that lexical differences between bilingual and monolinguals are persistent. This is not surprising, as vocabulary is a 'moving target' that continues to grow also in monolingual children and does not have a point at which it is completed (Paradis et al., 2010). Previous research that found similarly sized vocabularies across bilingual and monolingual children compared the children's conceptual or total vocabularies, which are vocabulary measures that include *both* languages of bilingual children (Hoff et al., 2012; Pearson et al., 1997). Other research that found no bilingual vocabulary delays compared bilingual children to monolingual norms (Pearson et al., 1993), instead of a control group, as was done in the present study. The normal range of variation is usually wide which reduces the likelihood of finding statistically significant differences between bilinguals and monolinguals (Bialystok, 2001), explaining discrepancies between studies that compare bilingual children with monolingual norms versus a monolingual control group.

In the present study, both the bilingual and monolingual children were highly accurate at narrative comprehension, indicating that the comprehension questions were relatively easy for 5- to 8-year-old children, in particular for the monolinguals over the full age range (5–8 years), and bilinguals from time 2 onwards (age 6–8 years). Listening comprehension scores were higher than generated story comprehension scores. High accuracies on MAIN comprehension questions are not atypical. Lindgren (2019) tested 17 Swedish monolingual children with MAIN at age 4;4 (time 1), 5;10 (time 2) and 7;4 (time 3) and observed a steep increase between times 1 and 2, but no further improvement between times 2 and 3. The lack of growth suggests a plateau effect that may stem from a lack of possibilities to grow, as children had, on average, between 80% and 90% correct at times 2 and 3. Rodina (2017) used MAIN with a sample of 4- to 6-year-old children (Russian-Norwegian bilingual, Russian and Norwegian monolingual), and found that the children's accuracies ranged between 7 and 10 points on a 10-point scale. Roch, Florit, and Levorato (2016) report an average of 6.5 questions (out of 9) correct for generated story comprehension and an average of 8.5 questions (out of 9) correctly answered questions after retelling a story in Italian by Italian-English children who are on average 5 years old. Similar accuracies are reported for generated story comprehension of bilingual Swedish-English 5-year old children (Bohnacker, 2016), and generated story comprehension and retelling of bilingual Polish-English (Otwinowska et al., 2018) 5-year old children.

In our study, as in other studies (Maviş et al., 2016; Otwinowska et al., 2018), the children were more accurate at listening comprehension, and answering the questions about a story told by someone else, than answering questions after story generation. This does not necessarily show that the children understand their own stories less well than those told by others. The stories that the children generate themselves can deviate from the 'expected' story. As a result, the pre-formulated comprehension questions may not always match the children's story, resulting in lower scores. When using pre-formulated comprehension questions, listening comprehension may thus be a better measure of narrative comprehension than comprehension questions that target a story that the children generated themselves. Interestingly, the children performed relatively well on the inferencing question about their own story (Appendix, Table 3), suggesting that they are able to draw correct inferences based on their own story.

At all three times, bilingual and monolingual children did not differ in listening comprehension, and the same was found at times 2 and 3 for generated story comprehension. Overall similar performance across the two groups was confirmed by more fine-grained comparisons of children's understanding of goals, internal states or inferences, and is in line with previous research (Boerma et al., 2016; Hipfner-Boucher et al., 2015; Rodina, 2017). Some caution is required, however, as ceiling performance may have masked better performance of the monolinguals, in particular for listening comprehension where scores were higher compared to generated story comprehension. As regards generated story comprehension, we found that at time 1, the monolingual children were more accurate than the bilingual children. In order to verify whether differences in nonverbal intelligence and age between the bilinguals and monolinguals could be responsible for the observed difference at time 1 on generated story comprehension, we compared subgroups of bilinguals and monolinguals matched on age and nonverbal intelligence suggests. This resulted in a p-value that just exceeded the α decision level, suggesting no difference. Note, however, that the combined effect of a smaller sample, use of non-parametric tests and application of a Bonferroni correction leads to an elevated risk of a type 2 error. It seems safe to conclude that there is some evidence for a difference between bilinguals and monolinguals on generated story comprehension, but the difference is small and disappears as a function of development.

4.2 Lexical and narrative comprehension: Variation within a bilingual sample

The Tarifit-Dutch and Turkish-Dutch groups did not differ in Dutch lexical comprehension. Whereas in the Turkish-Dutch group, input at home showed moderate to strong correlations with Dutch lexical comprehension, most correlations between input at home and lexical comprehension were close to zero in the Tarifit-Dutch group. Moreover, in the Turkish-Dutch group, lexical comprehension was positively correlated with SES. The relationship between SES and lexical comprehension was, in part, mediated by Dutch use at home, revealing that better lexical comprehension of Turkish-Dutch children with more highly educated parents is partly explained by the observation that more highly educated Turkish-Dutch parents use more Dutch at home when interacting with their children. Prevoo and colleagues (2013) found a similar pattern in this group of children. In their study, particularly Dutch reading activities, which were relatively more undertaken by the more highly educated parents, contributed to the mediation effect. In the Tarifit-Dutch group, SES did not correlate with Dutch lexical comprehension.

As regards listening comprehension, both bilingual groups showed the same highly accurate performance. The Turkish-Dutch group outperformed the Tarifit-Dutch group on generated story comprehension at time 2, but at times 1 and 3 no between-group difference emerged. Analyses with matched subgroups (age, nonverbal intelligence) confirmed the between-group difference at time 2, suggesting a robust effect. At first glance, the presence of a difference at time 2, and not at times 1 and 3, may be hard to interpret. Descriptive data suggest the same pattern at all three times, however, and given the small size of the subgroups, the absence of significance at times 1 and 3 may stem from a lack of power.

In the Tarifit-Dutch group, one relation with input reached significance after Bonferroni correction: the relation between home language richness and listening comprehension time 1. The language richness score is an aggregate score that measures the frequency of language use across several activities: reading (books, magazines, comics, newspapers), watching movies (television, cinema), story-telling, interacting with peers and friends. As these activities all involve understanding language beyond the level of individual words and sentences, it is not surprising that language use during these activities would support children's narrative comprehension. It is more striking that this relationship reached significance even though the home language richness score in the Tarifit-Dutch group is low, which could be related to a lack of resources in this language. Other relations did not reach significance but were moderately strong and potentially meaningful: Tarifit at home was positively correlated with narrative comprehension and Dutch use and richness correlated negatively with narrative comprehension. The observation that the relationships with Tarifit and Dutch are in opposite directions suggests that the type of input in these two languages is different and that, in the home environment, Tarifit input is more helpful for children's narrative comprehension than Dutch input. In the Turkish-Dutch group, input at home was not significantly correlated with narrative comprehension, but again there were several moderately strong correlations. The pattern that emerged in the Turkish-Dutch showed that Dutch use and richness were positively correlated with narrative comprehension, while home language use was negatively related with narrative comprehension.

The patterns in the Tarifit-Dutch group suggest that experiences in the home language can be transferred to narrative comprehension in the majority language (Pearson, 2002; Uccelli & Páez, 2007), whereas patterns in Turkish-Dutch group suggest that SES and Dutch input at home are related to Dutch lexicon and narration (Pearson, 2002; Prevoo et al., 2012). Relations between home language use are not found for lexicon, as expected, confirming that transfer is less important for bilingual children's lexical development in the majority language than for their narrative development. The observation that SES is correlated with lexical comprehension and not with narrative comprehension in our study may reflect the pattern found by Pearson (2002) who reports stronger relationships between SES and narrative microstructure, which includes lexical knowledge, than between SES and narrative macrostructure.

Intriguingly, across Tarifit-Dutch and Turkish-Dutch children, different patterns emerged. The results suggest that input differs across these two bilingual populations, in line with the expectations (Scheele et al., 2010), but also indicate that SES and language richness are differently related to Dutch language outcomes in the two groups. Parental education – the measure we used for SES – is an often-used measure in monolingual populations (Bornstein et al. 2003; Ensminger & Fothergill, 2003). However, differences in educational opportunities may imply that parental level of education does not index the same across cultures and countries. Within the scope of our research, we are unable to fully interpret the differential correlations that emerged between input and Dutch narrative comprehension. The reason is that the parent questionnaire we administered (PaBiQ, Tuller, 2015) provides information on language use and richness, but it does not provide insight into more detailed qualitative aspects of input at home such as richness of interaction and conversational style. Such factors are relevant for the narrative development of children growing up in low-SES families (Reese, Leyva, Sparks, & Grolnick, 2010), and could vary across language and cultures, and between the parents' first language (home language) versus second language (Dutch). The questions what parental education means across countries and cultures and how language practices at home vary across bilingual populations require further, in-depth research.

The home environment makes up only a part of the total input that children receive, raising the question whether differences in input *outside* the home environment can explain the differential patterns found for the two bilingual subgroups. The PaBiQ does not provide specific information on sources of Dutch input outside the home environment. However, for all children in the study, the school environment is a major input source, as children in the Netherlands spend, on average, more than 20 hours a week at school from age 4 years. We do not expect that exposure at school explains the variation, as the school situation (including access to early preschool and care facilities) is the same for the Turkish-Dutch and Tarifit-Dutch children and children in both groups were recruited from the same schools.

4.3 Clinical and educational implications

The results of this study have two important implications. First, they confirm that lexical measures are not well suited for identifying a Developmental Language Disorder in bilingual children, as bilingual children with typical language development are outperformed by their monolingual peers (Engel de Abreu et al., 2013).

Narrative macrostructure measures, in contrast, may be more suited as these show limited effects of bilingualism, but are sensitive to DLD (Boerma et al., 2016). Note that from a clinical perspective, the high accuracies that emerged in the current study are unproblematic, as long as children with DLD are not at ceiling. Previous research with MAIN shows that at time 1, 5 and 6-year old monolingual Dutch children with DLD score, on average, 8.1 ($SD = 1.8$) on listening comprehension (out of 10 questions) (Blom & Boerma, 2016). The same study shows that one year later, the children with DLD have improved and score 8.9 ($SD = 1.2$). Moreover, only at time 1, typically developing controls outperformed the DLD group, suggesting that at later ages, listening comprehension as measured by MAIN is not optimal for identifying DLD in bilingual children. Second, as lexical knowledge in the majority language is related to educational achievements (Snow et al., 1998; Dickinson & Tabors, 2001; Oller & Pearson, 2002; August & Shanahan, 2006), the persistent lexical gap is a risk factor that may threaten equal educational opportunities, in particular in the case of low-SES bilingual children because effects of SES and distributed input may add up.

4.4 Limitations of the study and future research

Several comparisons and correlations in the present study did not reach significance. This may signal that differences or relationships do not exist, but other explanations should be considered as well. Ceiling performance may have limited the possibility of finding between-group differences on narrative comprehension. The sample in the study was relatively small – in particular in the analyses focused on bilingual subgroups – and further reduced due to missing values; a lack of power could be a cause of several null results. We suggest future cross-language-domain longitudinal research with larger samples that allow for growth curve analyses. In previous research, we found that home language use and richness measured with the PaBiQ show significant and moderate to strong correlations with home language vocabulary (Blom, 2019). Moreover, differences between Turkish-Dutch and Tarifit-Dutch in language use at home pattern as expected (Scheele et al. 2010), suggesting that the PaBiQ is an adequate general measure of input at home. The PaBiQ is also limited, however, and does not provide insight into qualitative aspects of interactions, as pointed out earlier. Future research could make use of more detailed parent questionnaires and/or home observations across different bilingual populations. The use of parental education as an index of SES also requires further research, specifically in relation to different bilingual populations and different countries where bilinguals in the target groups migrated from.

4.5 Conclusions

Although bilingual Tarifit-Dutch and Turkish-Dutch children are outperformed by monolingual Dutch children on a single-language measure of lexical comprehension, their outcomes on a single-language measure of narrative comprehension did not differ. The first conclusion is therefore that bilingual children may lag behind in the majority language in one language domain (lexical comprehension), but that this does not generalize to another language domain (narrative comprehension). A second conclusion is that home language input, and particularly use of the home language during different activities (reading, watching movies, telling stories, conversations with friends) may support bilingual migrant children's narrative comprehension in the majority language. In the present study, this effect was, however, only found in the Tarifit-Dutch group and not in the Turkish-Dutch group. In contrast, effects of SES were related to Dutch lexical comprehension in the Turkish-Dutch group only. This brings us to the third conclusion that relations between factors in the home environment of migrant children and majority language development may hold for one bilingual group but not for the other. This finding emphasizes the importance of research that considers the home environment of bilingual children from different language and cultural backgrounds.

Funding

Funded by: Netherlands Organization for Scientific Research.
Award ID: 016.124.369.
Award recipient: Elma Blom.
Statement: This work is part of the research program 'Cognitive development in the context of emerging bilingualism: Cultural minority children in the Netherlands' which is financed by a VIDI-grant awarded to Elma Blom by the Netherlands Organization for Scientific Research (NWO; grant number 016.124.369).

References

August, D., & Shanahan, T. (2006). *Report of the National Literacy Panel on Language-Minority Children and Youth. Developing literacy in second-language learners*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Brown, P., Egorov, D., Mailhammer, R., Grant, A., & Holman, E. W. (2009). Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology*, 13(1), 167–179. https://doi.org/10.1515/LITY.2009.009

Berman, R., & Slobin, D. (1994). *Relating events in narrative: A crosslinguistic developmental study*. New York, NY: Psychology Press.

Bialystok, E. (2001). *Bilingualism in Development: Language, literacy, and cognition*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511605963

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(4), 525–531. https://doi.org/10.1017/S1366728909990423

Blom, E. (2019). Domain-general cognitive ability predicts bilingual children's receptive vocabulary in the majority language. *Language Learning*, 69(2), 292–322. https://doi.org/10.1111/lang.12333

Blom, E., & Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure? *Research in Developmental Disabilities*, 55, 301–311. https://doi.org/10.1016/j.ridd.2016.05.001

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

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Bornstein, M., Hahn, C., Suwalsky, J., & Haynes, O. (2003). Socioeconomic status, parenting, and child development: The Hollingshead four-factor index of social status and the socioeconomic index of occupations. In M. Bornstein & R. Bradley (Eds), *Socioeconomic status, parenting, and child development* (pp. 29–82). Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Abingdon, UK: Routledge.

COST Action IS0804 (2011). Questionnaire for Parents of Bilingual Children (PaBiQ). Retrieved from <http://www.bi-sli.org> (22 June, 2020).

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49(2), 222–251. https://doi.org/10.3102/00346543049002222

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70–89). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620652.006

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853596773

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, 28(3), 411–424. https://doi.org/10.1017/S0142716407070221

De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2013). A bilingual-monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, 35(6), 1189–1211. https://doi.org/10.1017/S0142716412000744

Dickinson, D. K., & Tabors, P. O. (2001). *Beginning literacy with language: Young children learning at home and school*. Baltimore, MD: Paul H. Brookes.

Dixon, L. Q., Wu, S., & Daraghmeh, A. (2012). Profiles in bilingualism: Factors influencing kindergartners' language proficiency. *Early Childhood Education Journal*, 40(1), 25–34. https://doi.org/10.1007/s10643-011-0491-8

Engel de Abreu, P., Baldassi, M., Puglesi, M., & Befi-Lopez, D. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: Testing minority-language children with an immigrant background. *Journal of Speech, Language, and Hearing Research*, 56(2), 630–642. https://doi.org/10.1044/1092-4388(2012/12-0079)

Ensminger, M., & Fothergill, K. (2003). A decade of measuring SES: What it tells us and where to go from here. In M. Bornstein & R. Bradley (Eds.), *Socioeconomic status, parenting and child development* (pp. 13–28). Mahwah, NJ: Lawrence Erlbaum Associates.

Extra, G., Aarts, R., Van der Avoid, T., Broeder, P., & Yağmur, K. (2001). *Meertaligheid in Den Haag: De status van allochtone talen thuis en op school*. Amsterdam: European Cultural Foundation.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools*, 35(2), 155–168. https://doi.org/10.1044/0161-1461(2004/016)

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics, 56*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Gagarina, N., Klop, D., Tsimpli, I-M., & Walters, J. (2016). Narrative abilities in bilingual children. *Applied Psycholinguistics*, 37(1), 11–17. https://doi.org/10.1017/S0142716415000399

Gathercole, V. C. M., & Hoff, E. (2007). Input and the acquisition of language: Three questions. In E. Hoff & M. Shatz (Eds.), *The handbook of language development* (pp. 107–127). Malden, MA: Blackwell.

Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, 12(2), 213–237. https://doi.org/10.1017/S1366728909004015

Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority L1 children learning English as a L2. *Applied Psycholinguistics*, 29(1), 1–25. https://doi.org/10.1017/S014271640808003X

Grosjean, F., & Li, P. (2012). *The psycholinguistics of bilingualism*. New York, NY: John Wiley & Sons.

Gutiérrez-Clellen, V. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education*, 13(2), 175–197. https://doi.org/10.1016/S0898-5898(01)00061-4

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, 29(4), 715–733. https://doi.org/10.1016/j.ecresq.2014.05.008

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.

Heilmann, J., Miller, J. F., & Nockerts, A. (2010). Sensitivity of narrative organization measures using narrative retells produced by young school-age children. *Language Testing*, 27(4), 603–626. https://doi.org/10.1177/0265532209355669

Hipfner-Boucher, K., Milburn, T., Weitzman, E., Greenberg, J., Pelletier, J., & Girolametto, L. (2015). Narrative abilities in subgroups of English language learners and monolingual peers. *International Journal of Bilingualism*, 19(6), 1–14. https://doi.org/10.1177/1367006914534330

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, 26(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E., & Core, C. (2013). Input and language development in bilingually developing children. *Seminars in Speech and Language*, 34(4), 215–226. https://doi.org/10.1055/s-0033-1353448

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1–27. https://doi.org/10.1017/S0305000910000759
Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., & Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology*, 43(5), 1062–1083. https://doi.org/10.1037/0012-1649.43.5.1062
Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of preschool children with typical and impaired language. *Bilingualism: Language and Cognition*, 15(1), 58–74. https://doi.org/10.1017/S1366728911000538
Lindgren, J. (2019). Comprehension and production of narrative macrostructure in Swedish: A longitudinal study from age 4 to 7. *First Language*, 39(4), 412–432. https://doi.org/10.1177/0142723719844089
Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605–621. https://doi.org/10.1080/20445911.2013.795574
MacLeod, A. A., Fabiano-Smith, L., Boegner-Pagé, S., & Fontolliet, S. (2013). Simultaneous bilingual language acquisition: The role of parental input on receptive vocabulary development. *Child Language Teaching and Therapy*, 29(1), 131–142. https://doi.org/10.1177/0265659012466862
Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, 37(1), 69–89. https://doi.org/10.1017/S0142716415000429
Nippold, M. A., Ward-Lonergan, J. M., & Fanning, J. L. (2005). Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic, and pragmatic development. *Language, Speech, and Hearing Services in Schools*, 36(2), 125–138. https://doi.org/10.1044/0161-1461(2005/012)
OECD (2018). Language barriers and the resilience of students with an immigrant background. In *The resilience of students with an immigrant background. Factors that shape well-being* (pp. 117–147). Paris: OECD Publishing. https://doi.org/10.1787/9789264292093-12-en
Oller, D. K., & Pearson, B. Z. (2002). Assessing the effects of bilingualism: A background. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 3–21). Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853595721-002
Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2018.1434124
Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, 1(3), 213–237. https://doi.org/10.1075/lab.1.3.01par
Paradis, J., Emmerzael, K., & Sorenson Duncan, T. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, 43(6), 474–497. https://doi.org/10.1016/j.jcomdis.2010.01.002
Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds), *Language and literacy in bilingual children* (pp. 135–174). Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853595721-008
Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43(1), 93–120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language*, 22(2), 345–368. https://doi.org/10.1017/S030500090000982X
Pearson, B. Z., Fernández, S. C., Ledeweg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18(1), 41–58. https://doi.org/10.1017/S0142716400009863
Prevoo, M., Malda, M., Mesman, J., Emmen, R., Yeniad, N., van IJzendoorn, M., & Linting, M. (2013). Predicting ethnic minority children's vocabulary from socioeconomic status, maternal language and home reading input: Different pathways for host and ethnic language. *Journal of Child Language*, 41(5), 963–984. https://doi.org/10.1017/S0305000913000299
Prevoo, M., Malda, Emmen R., Yeniad, N., & Mesman, J. (2015). A context-dependent view on the linguistic Interdependence Hypothesis: Language use and SES as potential moderators. *Language Learning*, 65(2), 449–469. https://doi.org/10.1111/lang.12099
Quiroz, B. G., Snow, C. E., & Zhao, J. (2010). Vocabulary skills of Spanish–English bilinguals: Impact of mother–child language interactions and home language and literacy support. *International Journal of Bilingualism*, 14(4), 379–399. https://doi.org/10.1177/1367006910370919
Reese, E., Leyva, D., Sparks, A., & Grolnick, W. (2010). Maternal elaborative reminiscing increases low-income children's narrative skills relative to dialogic reading. *Early Education and Development*, 21(3), 318–342. https://doi.org/10.1080/10409289.2010.481552
Richardson, J. T. E. (2011). Eta squared and partial eta squared as measurements of effect size in educational research. *Educational Research Review*, 6(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001
Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37(1), 49–67. https://doi.org/10.1017/S0142716415000417
Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian-Russian children. *International Journal of Bilingualism*, 21(5), 617–635. https://doi.org/10.1177/1367006916643528
Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, 323(5916), 951–953. https://doi.org/10.1126/science.1167025
Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31(1), 117–140. https://doi.org/10.1017/S0142716409990191
Schelletter, C. (2002). The effect of form similarity on bilingual children's lexical development. *Bilingualism: Language and Cognition*, 5(2), 93–107. https://doi.org/10.1017/S1366728902000214
Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Amsterdam: Harcourt Test.
Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, 30(2), 315–337. https://doi.org/10.1017/S0142716409090134
Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 301–330). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-013

Uccelli, P., & Páez, M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, 38(3), 225–236. https://doi.org/10.1044/0161-1461(2007/024)
Unsworth, S. (2013). Current issues in multilingual first language acquisition. *Annual Review of Applied Linguistics*, 33, 21–50. https://doi.org/10.1017/S0267190513000044
Verhoeven, L. (2007). Early bilingualism, language transfer, and phonological awareness. *Applied Psycholinguistics*, 28(3), 425–439. https://doi.org/10.1017/S0142716407070233
Wang, M., Park, Y., & Lee, K. R. (2006). Korean-English biliteracy acquisition: Cross language and orthography transfer. *Journal of Educational Psychology*, 98(1), 148–158. https://doi.org/10.1037/0022-0663.98.1.148
Wang, M., Perfetti, C. A., & Liu, Y. (2005). Chinese-English biliteracy acquisition: Cross language and writing system transfer. *Cognition*, 97(1), 67–88. https://doi.org/10.1016/j.cognition.2004.10.001
Westby, C. E. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and reading disabilities* (pp. 157–232). Boston, MA: Allyn & Bacon.
Wechsler, D., & Naglieri, J. A. (2008). *WNV NL. Wechsler Nonverbal Scale of Ability* (Nederlandstalige bewerking). Amsterdam: Pearson.
Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–2152. https://doi.org/10.1177/0956797613488145
Wichmann, S., Holman, E. W., & Brown, C. H. (Eds.). (2016). *The ASJP Database, version 17*. <http://asjp.clld.org/> (22 June, 2020).

Appendix 1. Listening comprehension, sum of points per question (number of children who received this question)

	Time 1		Time 2		Time 3	
	Bilingual	**Monolingual**	**Bilingual**	**Monolingual**	**Bilingual**	**Monolingual**
1	62 (66)	44 (44)	64 (65)	45 (45)	61 (61)	44 (44)
2	64 (66)	45 (45)	64 (65)	44 (45)	61 (61)	44 (44)
3	55 (67)	43 (45)	59 (66)	41 (44)	56 (60)	40 (44)
4	62 (67)	44 (45)	64 (66)	44 (45)	60 (61)	43 (43)
5	65 (67)	45 (45)	65 (66)	45 (45)	61 (61)	44 (44)
6	64 (67)	45 (45)	64 (66)	44 (45)	58 (60)	43 (43)
7	66 (67)	45 (45)	64 (66)	44 (45)	59 (61)	42 (43)
8	50 (66)	40 (44)	57 (66)	43 (45)	58 (61)	40 (44)
9	44 (66)	39 (44)	55 (66)	43 (45)	56 (61)	37 (44)
10	39 (67)	23 (45)	54 (66)	24 (44)	41 (60)	19 (44)

Note. 1 = goal episode 1; 2 = internal state as reaction episode 1; 3 = explanation internal state as reaction episode 1; 4 = goal episode 2; 5 = internal state as reaction episode 2; 6 = explanation internal state as reaction episode 2; 7 = goal episode 3; 8 = internal state as reaction episode 3; 9 = explanation internal state as reaction episode 3; 10 = inference based on whole story.

Appendix 2. Generated story comprehension (sum of points per question), sum of points per question (number of children who received this question)

	Time 1		Time 2		Time 3	
	Bilingual	**Monolingual**	**Bilingual**	**Monolingual**	**Bilingual**	**Monolingual**
1	48 (63)	41 (45)	57 (66)	43 (45)	60 (61)	43 (43)
2	47 (66)	37 (45)	57 (66)	37 (45)	50 (61)	40 (44)
3	35 (65)	35 (45)	42 (63)	34 (45)	45 (61)	36 (44)
4	62 (64)	43 (45)	63 (66)	45 (45)	61 (61)	43 (44)
5	60 (66)	45 (45)	63 (66)	44 (45)	61 (61)	44 (44)
6	51 (64)	44 (45)	61 (66)	43 (45)	59 (61)	42 (44)
7	46 (61)	41 (45)	57 (66)	42 (44)	55 (60)	40 (44)
8	28 (62)	29 (45)	42 (66)	31 (44)	48 (61)	35 (44)
9	20 (62)	21 (45)	36 (66)	27 (44)	43 (61)	34 (44)
10	50 (64)	39 (45)	53 (65)	41 (44)	59 (61)	40 (43)

Note. 1 = goal episode 1; 2 = internal state as reaction episode 1; 3 = explanation internal state as reaction episode 1; 4 = goal episode 2; 5 = internal state as reaction episode 2; 6 = explanation internal state as reaction episode 2; 7 = goal episode 3; 8 = internal state as reaction episode 3; 9 = explanation internal state as reaction episode 3; 10 = inference based on whole story.

Why do you think the boy would be unhappy if he saw what the cat was eating?

Comprehension of German narratives in Russian- and Turkish-German bilingual children

Natalia Gagarina[1,2], Nathalie Topaj[1] and Natalie Sürmeli[1]
[1]Leibniz-Centre General Linguistics (ZAS) / [2]Uppsala University

The present study traces longitudinal trajectories of narrative comprehension in L2 German in 57 Russian- and Turkish-German bilingual children, for two age groups with a mean starting age of 3;6 and 4;3, respectively. The children answered comprehension questions from the *Multilingual Assessment Instrument for Narratives* in their L2 German. They were tested three times with a one-year interval in between, so that by the last time children had a mean age of 5;6 and 6;2, respectively. Comprehension questions targeted two types of elements of macrostructure, goals and internal states, in two modes, telling and model story. Results showed a significant improvement of narrative comprehension over time in both age groups. The steepest slope of development was found between three and five years. By age five, the children were able to answer the majority of comprehension questions, whilst after that, a plateau in the development was observed. Generally, goals were easier to comprehend than internal states, at all test times, for all groups and elicitation modes. The study enriches our knowledge of L2 bilingual trajectories of the comprehension of goals and internal states by showing that these two components of macrostructure require different inferencing skills and that goals are cognitively easier to infer as compared to internal states.

Keywords: narrative comprehension, macrostructure, bilingual children, German, MAIN

1. Introduction

Narrative comprehension is an indispensable part of narrative competence. It involves, among other things, the understanding and interpretation of visual and/or oral stimuli (depending on the task), i.e. the mental construction of a basic story organization on a macrostructure level (Bohnacker, 2016; Gagarina et al., 2012;

https://doi.org/10.1075/sibil.61.08gag

Stein & Glenn, 1979; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994; among others). It also includes the ability to infer protagonists' feelings and thoughts, i.e. to literally get into their mind, which is often called *theory of mind* (e.g., De Cat, 2008; Leslie, 1987; Tomasello, 2003; Wellman, Cross, & Watson, 2001). Last but not least, memory capacity, information processing, and other cognitive abilities are also required for the development of narrative comprehension (Bel, Ortells, & Morgan, 2015). Thus, narrative comprehension includes a whole complex of linguistic and cognitive skills which gradually develop with age before reaching the adult target level.

Language comprehension precedes production in general language development (Hendriks & Koster, 2010) and researchers mainly investigate comprehension of the lexicon and grammar, following its development throughout childhood. Much less attention has been paid to the comprehension of narratives, e.g., stories expressed by means of pictorial and/or oral stimuli. Only recently (with some exceptions) have researchers taken up this interest (see, e.g., Bohnacker, 2016; Gagarina et al., 2012; Lindgren, 2018; Maviş, Tunçer, & Gagarina, 2016; Stein & Glenn, 1979; Trabasso & Rodkin, 1994; among others). Given the importance of narrative comprehension in educational and professional contexts, it is surprising that comprehension of content in oral texts has received so little attention so far.

Development of narrative skills follows the acquisition of basic lexical, syntactic, and morphological categories of a language (cf. Bamberg, 1994; Berman & Slobin, 1994; Hickmann, 2003; among others). While narrative production can be assessed by the onset of active narrative development starting at around age four (e.g., Berman & Slobin, 1994; Berman, 2009; Stein, 1988), assessing narrative skills through comprehension tasks is possible with younger children who still have quite restricted productive abilities. This applies not only to monolingual but also to bilingual children, who might acquire their languages at different ages.

Although research using data collected from bilingual children is still underrepresented, bilingual children have been found to acquire macrostructure in the same manner as monolingual children (cf. Akinci, Jisa, & Kern, 2001; Fiestas & Peña, 2004; Gagarina et al., 2015; Kunnari, Välimaa, & Laukkanen-Nevala, 2016; among others). Moreover, bilingual children who acquire two languages sequentially may transfer their narrative skills from one language to another to some degree: "successive bilinguals catch up more quickly for story structure because it is a cognitive-linguistic interface skill that can be shared with the first language (Paradis et al., 2011; Paradis & Schneider, 2008; Pearson, 2002)" (Bohnacker, 2016: 23–24).

Research on comprehension more frequently addresses the comprehension of written texts (reading comprehension) (cf. Wagner, Schatschneider, & Phythian-Sence, 2009). Much less is known about the comprehension of visually presented narratives or orally and visually presented narratives examined via answers to comprehension questions (Stein & Glenn, 1979; Trabasso & Nickels, 1992). One study that is relevant for our research by Stein & Glenn (1979) examined answers to comprehension questions in 6- and 10-year-old monolingual children. Their results show good performance in comprehension already at age six, whereas production lags behind comprehension. The focus of the comprehension part of that study was on goals and internal states, components of the macrostructure which require inferencing. Much later, Trabasso and Nickels (1992) found that monolingual English-speaking children showed comprehension of protagonists' goals already at age four.

In the following, we will concentrate on studies which have assessed narrative comprehension by using the *Multilingual Assessment Instrument for Narratives* (MAIN) which is part of *Language Impairment Testing in Multilingual Settings* (LITMUS, Gagarina et al., 2012, 2015). The recent development of MAIN for narrative comprehension and its application across various languages and populations allows for comparable methods of elicitation and analysis of data with regard to the comprehension of the main content of narratives. How is it possible to assess comprehension? Within the framework of MAIN, participants tell or retell picture-based stories and then answer comprehension questions. The comprehension questions target the higher-order organization of narratives and address two components of story structure – goals and internal states. This method has already been applied to various languages in bilingual and monolingual children (Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Bohnacker, 2016; Gagarina, 2016; Kapalková, Polišenská, Marková, & Fenton, 2016; Kunnari, Välimaa, & Laukkanen-Nevala, 2016; Lindgren, 2018; Maviş, Tunçer, & Gagarina, 2016; Otwinowska, Mieszkowska, Białecka-Pikul, Opacki, & Haman, 2018; Roch, Florit, & Levorato, 2016; Tsimpli, Peristeri, & Andreou, 2016; among others).

For example, Bohnacker (2016) investigated narrative production and comprehension in Swedish-English bilinguals in two age groups: 5- and 6–7-year-olds. Children displayed a higher level of comprehension than production and there was a significant development over time between the two age groups with regard to narrative production, but not comprehension. The results suggest that 5-year-old children already have a relatively high level of narrative comprehension while their production is still developing. The results were similar in both languages, indicating consistency of story structure across languages in the groups of bilinguals. At the same time, a more thorough qualitative analysis of narrative comprehension with regard to different story structure components showed that questions targeting internal states of protagonists as initiating events or reactions were more difficult for children than those targeting goals. One of the internal state questions requires the child to reflect on the whole story and to use theory of mind, which proved especially difficult. Bohnacker concluded that not only overall scores but also qualitative differences should be analysed in order to document individual and age differences in narrative production and comprehension of the story structure components.

In another study using MAIN to assess narrative production and comprehension, Lindgren (2018) investigated narrative competence in Swedish monolinguals and German-Swedish and Turkish-Swedish bilinguals aged 4 to 6. She found that, already at age 4, children were quite advanced in comprehension whereas their production of macrostructure was still rudimentary. Both skills developed with age. In production, however, children rarely verbalized all components of the story structure even at age 6, especially if they related to theory of mind. Interestingly, the results for bilingual children were different across languages. Turkish-Swedish bilinguals scored higher in Turkish or equally in both languages (depending on measures) whereas German-Swedish bilinguals were better in Swedish. At the same time, Turkish-Swedish bilinguals scored lower overall than German-Swedish ones. However, these results could be also connected to differences in group composition, as bilingual groups may be relatively heterogeneous in terms of their bilingual type, age of onset, chronological age, social background, etc. Several further studies investigated narrative comprehension and production in both languages of bilingual children, e.g., in Slovak-English 5–6-year-olds (Kapalková et al., 2016) as well as in Italian-English 5–6-year-olds and 6–7-year-olds (Roch et al., 2016). The authors found that in the first language pair the narrative comprehension was similar in both languages whereas in the second language pair it was better in L1 Italian in younger children and it did not differ between the languages in older children. In another language pair, Polish-English, 4- to 6-year-old bilingual children showed better story production and comprehension when they had first listened to the model story told by an examiner (Otwinowska et al., 2018). Here, the bilingual children did not significantly differ from Polish monolinguals.

For one of the language combinations included in the present study, Turkish-German, Maviş et al. (2016) examined and compared production and comprehension elicited in different modes in Turkish. Children were divided into three age groups, 2;11–3;11, 4;0–5;11, and 6;0–7;1. With regard to comprehension they found that children scored significantly better when comprehension questions were asked after listening to the model story than when the comprehension questions were asked after the children had told the story themselves ('telling', without having listened to the model story), especially in the youngest age group. In the model story mode, the story is told by an examiner to provide a model while children look at the pictures and the comprehension questions are asked immediately after; in the telling mode, the story is told by the children themselves on the basis of pictures shown to them. They also found effects of age (the older groups performed significantly better than the youngest group in both elicitation modes).

To sum up, the studies reviewed above showed higher performance in comprehension compared to production, and a development in the general comprehension score in children of preschool age. Although children achieved high scores in comprehension, they still were not able to answer some questions involving theory of mind very well. Only two studies (Bohnacker, 2016; Lindgren, 2018) have performed an in-depth examination of the comprehension of two components of macrostructure, goals and internal states, and they found the latter to be more difficult in comprehension.

Overall, the results are not completely uniform even when the same elicitation methods and comparable analyses have been reported, as in the case of studies using MAIN. The investigation of narrative competence in different language constellations and groups gives new insights into its acquisition and development in children and adults of different ages and backgrounds. Our study broadens the range of language combinations to children who acquire German as their L2 and have Russian or Turkish as their home language, and provides an in-depth analysis of the comprehension of two components of macrostructure – goals and internal states. Another novel aspect are our longitudinal trajectories of comprehension development in children of various ages over a period of two years.

In bilingual language acquisition, age of onset and length of exposure play a decisive role in the attainment of target language proficiency. This role is, however, not straightforward but interacts with various linguistic and environmental background factors. The interplay of age of onset, length of exposure, and various factors has been intensively discussed in the last decades. Increased length of exposure moderated by other factors has been shown to have a positive impact on lexical and grammatical development (Paradis, 2019; Schulz & Grimm, 2019). As the impact of age of onset and length of exposure in the development of narrative skills in L2 German has not yet been investigated, the present study aims to fill this gap.

2. Aims and research questions

Our study traces comprehension of the content of picture-based narratives in German by evaluating the responses to comprehension questions targeting macrostructure, in particular goals and internal states as components of story episodes. We examine the longitudinal development of narrative comprehension in 2- to 6-year-old Russian-German and Turkish-German bilinguals. Given that the age range of the sample at the first test time was between 2;10 and 4;7 (which includes a crucial period of language acquisition in general, and of narrative skills in particular) the sample has been divided into two age groups in order to reach higher homogeneity in terms of age and length of exposure and to achieve higher robustness of the findings. Specifically, we compare the developmental trajectories across three test times with a one-year interval in two age groups: younger than four (age range 2;10–3;11) and above four (age range 4;0–4;7) at the time of the first testing. These analyses allow not only for longitudinal but also for cross-sectional comparisons. The study is guided by the following research questions:

1. How does narrative comprehension in German develop in bilingual children across three test times in two groups with different ages but similar length of exposure to German?
2. Does the elicitation mode (model story versus telling) in any way impact comprehension of story structure?
3. Does comprehension of two types of macrostructure components, goals and internal states, significantly differ across age groups and elicitation modes?

3. Method

3.1 Participants

For the current investigation we analysed the data of 57 Russian-German and Turkish-German bilingual children who participated in a large longitudinal study with 160 children conducted in Berlin by the *Berlin Interdisciplinary Alliance for Multilingualism* (BIVEM). The BIVEM study addressed among other things the effects of different language support methods over a period of three years and examined the language development of children in each of their languages. Russian and Turkish are the most frequent languages among children with migrant backgrounds in Germany and are therefore of great interest for studies on bilingual children. Given that language support methods are not the focus of the present investigation and that the number of children for the current investigation is not very large, we combined all intervention groups. Before taking this step, we compared the results on narrative comprehension in different intervention groups in the investigated corpus and found no significant differences between them at any of the investigated test times. However, it cannot be excluded that interventions could play a role in narrative comprehension in larger groups, therefore, we keep it in mind for further investigations.

The data were collected at three test times, the interval between the test times being 11–12 months. As the age range at time1 was large, spanning 2;11–4;7, we divided the participants into two groups: age group 1 (AG1) including 33 children (mean age 42 months, age range 2;10–3;11) and age group 2 (AG2), including 24 children (mean age 51 months, age range 4;0–4;7). The group means and standard deviations (SD) are given in Table 1. Although the main study offers much more data, for the current investigation only those children have been considered who provided a full data set for the analysis of narrative comprehension at all test times (no instances of test discontinuation were included in the present sample).

We took into account several parameters in group composition, using information obtained from a parental questionnaire (based on the questionnaire used in the *Russian language proficiency test for multilingual children* (Gagarina, Klassert, & Topaj, 2010) and adapted to Turkish as well), which included information on the child's language and social background as well as that of the parents.

The age groups had a comparable length of exposure (LoE) of 27 months in AG1 and 30 months in AG2 to L2 German (language of environment); the age of onset (AoO) differs by six months: the mean AoO in AG1 is 15 months (*SD* = 12) and the mean AoO in AG2 is 21 (*SD* = 13). We follow the assumption that, with an AoO below three years, bilingual children acquire their German not as L2 learners but more like child L1 learners (bilingual first language acquisition, e.g., Meisel, 2008, 2011). At the same time, we cannot exclude that later AoO or shorter LoE (being complementary to each other) can influence the comprehension and production outcomes to a certain degree at different stages of language acquisition in our age range. Both age groups included children with either Russian or Turkish as the home language: 14 Turkish- and 19 Russian-speaking children in AG1, and 13 Turkish- and 11 Russian-speaking children in AG2. The number of boys and girls was also comparable in the age groups: 19 girls, 14 boys in AG1 and an equal number of 12 boys and 12 girls in AG2.

Table 1. Description of the sample by age group

Sample details	**Age group 1 (*N* = 33)**	**Age group 2 (*N* = 24)**
Age range in months (y;m)*	34–47 (2;10–3;11)	48–55 (4;0–4;7)
Age in months: mean (*SD*)*	42.2 (*3.5*)	50.8 (*1.8*)
Age of onset (AoO) in months: mean (*SD*)*	15 (*12*)	21 (*13*)
Length of exposure (LoE) to L2 German in months: mean (*SD*)*	27 (*12*)	30 (*14*)
Home language (L1) Turkish / Russian (N)	14 / 19	13 / 11
Socio-economic status (SES): mean (*SD*) [N**]	1.94 (*0.67*) [25]	1.94 (*0.44*) [18]
Girls / boys (N)	19 / 14	12 / 12
Intelligence quotient (IQ): mean (*SD*)	108.5 (*13.4*)	104.9 (*12.2*)

Notes: *N* = number of participants; *SD* = standard deviation
* at the first testing after the first intervention period (time1);
SES points are based on the parents' cumulative points for the highest degree of education: 1 – general certificate of secondary education, apprenticeship or equivalent, 2 – high school degree (German Abitur) or equivalent, 3 – university degree
** number of participants with available SES data.

The socio-economic status (SES) of the children was determined according to a slightly modified version of the procedure of the Berlin Senate (Bettge & Oberwöhrmann, 2017), based on the parents' cumulative highest degree of education: 1 – general certificate of secondary education, apprenticeship or equivalent, 2 – high school degree (German Abitur) or equivalent, and 3 – university degree. Children of both age groups had an equal SES mean of 1.94 based on the available data (the number of children with available data is given in square brackets). For 14 participants SES data are missing, as their parents did not provide this information in the questionnaire.

In addition, all participants were tested on their non-verbal intelligence to control for possible influences on language development. The IQ values were obtained on the basis of the Snijders-Oomen non-verbal intelligence test SON-R 2½-7 (Tellegen, Laros, & Petermann, 2007), normalized for the German population. Subtests on drawing conclusions (categorizing objects, finding analogies, and finding logical connections between situations), which have a separate IQ value, were performed during the first cycle of the project with all participating children. One child in the sample (AG1) had an IQ value of 77 at the time of testing, which is considered an exclusion criterion. However, his results for all language tests in German were within normal range, which is why the data of this child were not excluded from the analysis.

3.2 Materials

For the present study, we used MAIN (Gagarina et al., 2012, 2019). This tool was designed for testing the narrative abilities (production and comprehension) of monolingual and multilingual children from early childhood to 12 years old. We elicited narrative production and comprehension in two different modes: *model story* (the story is told by an examiner to provide a model) and *telling* (the story is told directly by a child). Four picture-based stories were used: Cat or Dog in the *model story mode*, and Baby Birds or Baby Goats in the *telling mode*. A description of the stories including the pictures can be found in MAIN (Gagarina et al., 2012, 2019).

The comprehension questions require inferencing of goals, feelings, and thoughts of protagonists and not just recollection of the content of the story. Of the ten questions constituting the set of comprehension questions in MAIN (D1–D10), three questions target goals of the story episodes (D1, D4, and D7), six questions target so-called internal states as reactions to the outcomes in different episodes and are asked in pairs (D2/D3, D5/D6, and D8/D9), and the last question (D10), which was not asked in the present study, relates to the whole story and not to a particular episode. Several questions targeting internal states as well as the last question involve inferencing and taking the perspective of different protagonists.

Questions asked about goals are formulated as follows: *Why does the dog jump?* (D1 / Dog story) or *Why does the bird bite the fox's tail?* (D7 / Baby Goats story). Questions asked about the internal states are formulated as follows: *How does the cat feel? / Why do you think that the cat is feeling angry/disappointed/hurt etc.?* (D2/D3, Cat story) or *Imagine that the dog sees the birds. How does the dog feel? / Why do you think that the dog feels good/fine/happy/satisfied etc.?* (D8/D9, Baby Birds story). In the case of internal states the second question in each pair is only asked if the child gives a correct answer to the previous question without an explanation.

3.3 Procedure

All tests were performed individually with each child by trained researchers and research assistants in a separate room in the kindergarten, providing a familiar and quiet setting. Stories were counterbalanced to testing within elicitation mode, language, and group in such a way that children of all groups always had a different story per language and mode. As mentioned above, Cat and Dog stories were used in the model story mode, Baby Birds and Baby Goats in the telling mode. This was done according to the randomization procedure described in MAIN. In our study, model story elicitation was performed at time1 and time2 test times, and telling elicitation at all three test times with all children.

The examiner[1] presented the stimulus materials according to the standardized MAIN procedure. First, the examiner showed the whole six-picture sequence to the child in such a way that only the child could see the pictures (non-joint attention mode) and asked him or her to look at the pictures. Then, the examiner explained that she was going to tell a story based on the presented pictures while only the child could see the pictures, and then would ask the child several questions afterwards, targeting his/her listening comprehension with visual support (model story elicitation mode). The examiner proceeded to tell the story, unfolding and revealing the pictures two by two at the corresponding points in the story. After conclusion of the model story part (including the comprehension questions), a second picture sequence was presented to the child, also in non-joint attention mode. The method for presenting these pictures was the same as with the model story, but this time the examiner asked the child to tell his/her own story (telling elicitation mode).

As previously mentioned, each story was followed by a set of comprehension questions. Children were asked to give answers to all questions in order to assess their narrative comprehension. Given that especially younger children (2–3-year-olds) might have difficulties with answering complex questions and may refuse to

1. All examiners were female, thus only the female form is used in the text when referring to them.

continue with testing, a discontinuation criterion was drawn up. If a child could not give an adequate answer to the first three questions or did not understand the questions at all, this part of the test was discontinued. If the answers were not correct but comprehensible, the examiner proceeded with asking all further questions. In a few cases of general refusal to answer the questions, this part of the test also had to be discontinued. Data of children who did not finish the comprehension question part of the test due to discontinuation were excluded from the analysis (the sample of 57 children did not contain these data).

The total score for questions D1–D9 constituted the basis of our analysis of the general development of narrative comprehension in bilingual children over time and in comparison between different elicitation modes. Separate scores for questions targeting goals and internal states were analysed in more detail, providing insight for a more specific analysis of the development of narrative comprehension. Question D10 was excluded from the total score as it was not presented to any of the children at test time1 and was presented to only some of the children at time2 (due to the absence of this question in an earlier version of MAIN used at the beginning of the project).

3.4 Statistical analysis

The statistical analysis was performed in R (version 3.3.3, R Core Team, 2017). We used multiple linear regression models in order to investigate the development of narrative comprehension over time in the age groups in relation to the elicitation mode and/or age group. At the same time, we took into account LoE at the time of the first testing and L1 as additional variables. Interactions between time and all mentioned independent variables were included to pinpoint their possible effects within the developmental trajectories. Single comparisons were performed with Welch t-tests or Wilcoxon tests for paired samples with adjusted p-values.

4. Results

In the first part of this section we provide the analyses based on the total scores for comprehension questions for each age group and elicitation mode at all test times. In the second part of this section we present analyses based on the scores for the comprehension of goals and internal states separately.

4.1 Total scores for comprehension questions

Table 2 provides an overview of all total scores in our analyses, i.e. correct responses for comprehension questions, including means, standard deviations (SD), and percentage for each age group, elicitation mode, and test time.

Table 2. Total scores for comprehension questions (out of 9 = 100%): Means, (*SD*), percentage by age group, elicitation mode, and test time

Elicitation mode	Age group	Mean age at Time1*	Time1	Time2	Time3
Telling	**AG1**	3;6	1.76 (*1.48*) 20%	5.27 (*1.89*) 59%	6.70 (*1.42*) 74%
	AG2	4;3	3.75 (*2.52*) 42%	6.42 (*1.89*) 71%	6.67 (*1.49*) 74%
Model	**AG1**	3;6	3.24 (*2.03*) 36%	7.00 (*1.64*) 78%	NA
story	**AG2**	4;3	4.25 (*2.71*) 47%	7.38 (*1.28*) 82%	

Note: NA = not available as model story was not performed at time3
* the mean age in each group increased by one year with each test time.

Figure 1 and Figure 2 demonstrate the distribution of scores (correct responses) for comprehension questions by age group and test time as well as their improvement over time for the telling and model story elicitation modes.

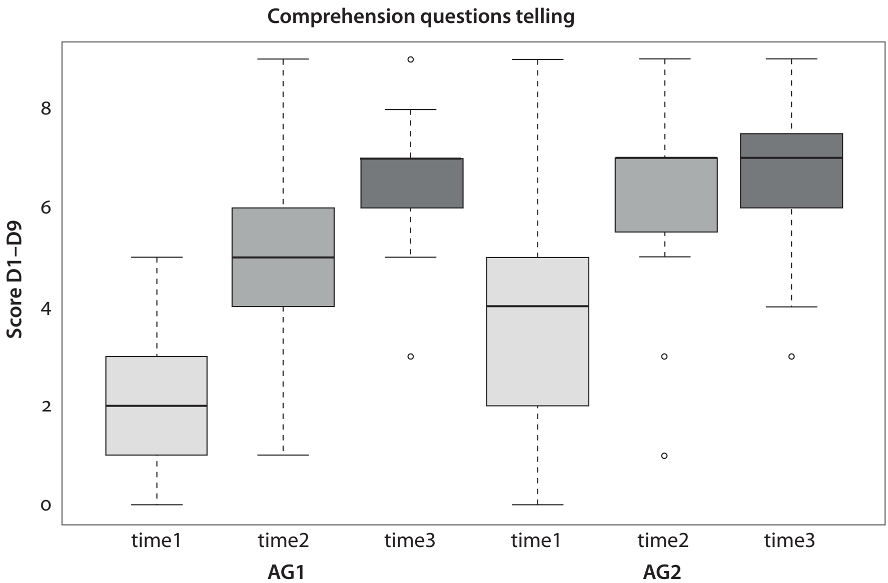

Figure 1. Distribution of correct responses to comprehension questions in the telling elicitation mode by age group and test time

The analysis of the total scores (correct responses) for comprehension questions in the **telling elicitation mode** showed that there is a significant development over time in both age groups: from only 1.76 points at time1 to 5.27 points at time2 and to 6.70 points at time3 in AG1 ($b = 3.26$, $SE = 1.19$, $t = 2.74$, $p < 0.01$ between time1 and time2; $b = 4.56$, $SE = 1.41$, $t = 3.25$, $p < 0.01$ between time1 and time3); from 3.75 points at time1 to 6.67 points at time 3 in AG2 ($b = 5.06$, $SE = 2.12$, $t = 2.38$, $p < 0.05$ between time1 and time3), according to the linear regression models presented in Table 3. In both age groups no other main effects were found, but interestingly there is a significant interaction between time3 and L1 Turkish in AG1 ($b = 1.62$, $SE = 0.80$, $t = 2.02$, $p < 0.05$), suggesting that in the younger age group the increase in the scores of Turkish-speaking children is higher than in the scores of Russian-speaking children between time1 and time3.

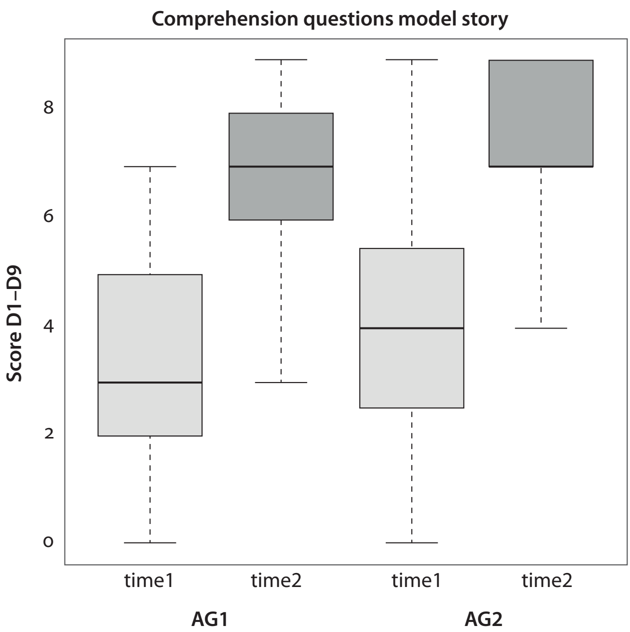

Figure 2. Distribution of correct responses to comprehension questions (D1–D9) in the model story elicitation mode: Distribution by age group and test time

The same analysis in the **model story elicitation mode** showed that in both age groups there is a significant development over time from 3.24 at time1 to 7.00 points at time2 in AG1 ($b = 3.73$, $SE = 1.39$, $t = 2.68$, $p < 0.01$) and from 4.25 to 7.38 points in AG2 for the same period of time ($b = 4.66$, $SE = 1.98$, $t = 2.36$, $p < 0.05$) with no effects of other investigated variables, according to the linear regression models presented in Table 3.

Table 3. Linear regression models for the analysis of development over time separately for age groups and elicitation modes

		Age group 1				Age group 2			
		b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Telling	(Intercept)	1.44	0.72	1.99	*	2.35	1.16	2.04	*
	Time2	3.26	1.19	2.74	**	3.03	1.88	1.62	
	Time3	4.56	1.41	3.25	**	5.06	2.12	2.38	*
	LoE	0.03	0.02	1.08		0.05	0.03	1.50	
	L1t	−0.87	0.57	−1.53		0.00	0.85	0.00	
	Time2:LoE	−0.02	0.03	−0.53		−0.02	0.04	−0.45	
	Time3:LoE	−0.02	0.03	−0.54		−0.06	0.04	−1.36	
	Time2:L1t	1.49	0.80	1.85		−0.18	1.20	−0.15	
	Time3:L1t	1.62	0.80	2.02	*	−0.14	1.19	−0.11	
Model story	(Intercept)	3.55	0.84	4.20	***	2.75	1.21	2.26	*
	Time2	3.73	1.39	2.68	**	4.66	1.98	2.36	*
	LoE	0.00	0.03	−0.10		0.03	0.03	1.04	
	L1t	−0.54	0.66	−0.81		0.89	0.89	1.00	
	Time2:LoE	−0.01	0.04	−0.30		−0.03	0.05	−0.67	
	Time2:L1t	1.20	0.94	1.28		−1.20	1.26	−0.95	

Notes: These models include fixed variables of time (time1, time2, time3), LoE, and L1; L1t refers to L1 Turkish as opposed to L1 Russian; significance levels * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

4.2 Comprehension questions with regard to the age group and elicitation mode

In the next step the comparison of developmental trajectories of the groups was analysed with regard to the effect of age group and its interaction with time (separately for the telling and model story modes) and with regard to the effect of elicitation mode and its interaction with time (separately for each age group). The developmental trajectories are shown in Figure 3.

With regard to the effect of age group and its interaction with time, the linear regression analyses showed that in the telling mode there are positive main effects of time ($b = 3.40$, $SE = 1.02$, $t = 3.32$, $p < 0.01$ for time2 and $b = 5.51$, $SE = 1.18$, $t = 4.68$, $p < 0.001$ for time3) and age group ($b = 1.95$, $SE = 0.49$, $t = 4.00$, $p < 0.001$ for AG2) and a negative interaction between time3 and AG2 ($b = -2.02$, $SE = 0.69$, $t = -2.94$, $p < 0.01$), as presented in Table 4. The results suggest that the developmental trajectories of the two age groups over the three test times are not similar: whereas the scores in AG1 continuously increase from time1 to time3, the scores in AG2 increase between time1 and time2, but not between time2 and time3. In other words, the increase is higher in AG1 than in AG2. In the model story mode

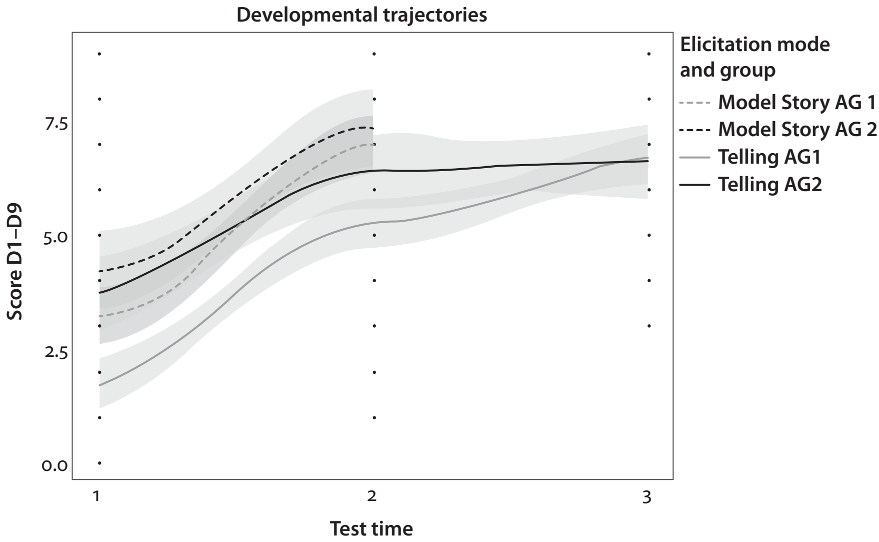

Figure 3. Developmental trajectories of both age groups based on the total scores for comprehension questions in the telling and model story elicitation modes

there is only a main effect of time ($b = 4.18$, $SE = 1.14$, $t = 3.67$, $p < 0.001$ at time2) but no effect of age group, according to the linear regression models presented in Table 4. No interaction was found between time and age group, suggesting that the developmental trajectories of the age groups are rather similar.

With regard to the effect of elicitation mode and its interaction with time, the analysis showed that there is a negative main effect of the elicitation mode telling for AG1 ($b = -1.48$, $SE = 0.44$, $t = -3.41$, $p < 0.001$), however not for AG2 (Table 5). At the same time, there is a positive effect of time for both age groups ($b = 3.62$, $SE = 0.98$, $t = 3.71$, $p < 0.001$ for AG1, $b = 4.08$, $SE = 1.47$, $t = 2.77$, $p < 0.01$ for AG2). Incidentally, the analysis for AG1 also showed an interaction between time2 and L1 ($b = 1.35$, $SE = 0.63$, $t = 2.15$, $p < 0.05$), indicating a higher increase in scores for Turkish-speaking children. There is no interaction between time and elicitation mode in either age group, suggesting a similar increase in scores in both age groups from time1 to time2.

Overall, the narrative comprehension of bilingual children of both age groups significantly improved in both elicitation modes over the time period of two years with regard to the total scores for comprehension questions.

Table 4. Linear regression models for the developmental trajectories in relation to the age group effect: Total score

	Telling				Model story			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
(Intercept)	1.06	0.63	1.68		2.89	0.70	4.13	***
Time2	3.40	1.02	3.32	**	4.18	1.14	3.67	***
Time3	5.51	1.18	4.68	***		NA		
AG2	1.95	0.49	4.00	***	0.96	0.54	1.78	
LoE	0.03	0.02	1.83		0.01	0.02	0.60	
L1t	−0.52	0.48	−1.10		0.03	0.53	0.05	
Time2:AG2	−0.90	0.69	−1.30		−0.60	0.77	−0.78	
Time3:AG2	−2.02	0.69	−2.94	**		NA		
Time2:LoE	−0.02	0.03	−0.59		−0.02	0.03	−0.60	
Time3:LoE	−0.04	0.03	−1.33			NA		
Time2:L1t	0.78	0.68	1.15		0.20	0.75	0.27	
Time3:L1t	0.92	0.68	1.37			NA		

Notes: These models include fixed variables of time (time1, time2, time3), LoE, and L1; L1t refers to L1 Turkish as opposed to L1 Russian; significance levels * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

Table 5. Linear regression models for developmental trajectories in relation to the elicitation mode effect: Total score

	Age group 1				Age group 2			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
(Intercept)	3.23	0.60	5.37	***	2.80	0.922	3.05	**
Time2	3.62	0.98	3.71	***	4.08	1.47	2.77	**
Mode_Telling	−1.48	0.44	−3.41	***	−0.50	0.63	−0.80	
LoE	0.01	0.02	0.62		0.04	0.02	1.73	
L1t	−0.71	0.44	−1.60		0.45	0.63	0.71	
Time2:Mode_Telling	−0.24	0.62	−0.40		−0.46	0.89	−0.52	
Time2:LoE	−0.01	0.03	−0.57		−0.03	0.03	−0.77	
Time2:L1t	1.35	0.63	2.15	*	−0.69	0.90	−0.77	

Notes: These models include fixed variables of time (time1, time2, time3), LoE, and L1; L1t refers to L1 Turkish as opposed to L1 Russian; significance levels * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

4.3 An in-depth analysis of goals and internal states

In the next step we analysed the scores for the comprehension of goals and internal states in the telling and model story elicitation modes. Table 6 shows the overall scores for the comprehension of goals or internal states in both elicitation modes, telling versus model story, presented by age group and test time.

Table 6. Scores for the comprehension of goals and internal states: Means, (*SD*), percentage by age group, elicitation mode, and test time

Elicitation mode	Age group	Time1	Time2	Time3
		Goals (out of 3 = 100%)		
Telling	**AG1**	0.85 (*0.76*) 28%	2.24 (*0.87*) 75%	2.76 (*0.50*) 92%
	AG2	1.75 (*1.11*) 58%	2.50 (*0.59*) 83%	2.71 (*0.55*) 90%
Model story	**AG1**	1.88 (*1.02*) 63%	2.79 (*0.55*) 93%	NA
	AG2	2.17 (*0.96*) 72%	2.88 (*0.34*) 96%	
		Internal states (out of 6 = 100%)		
Telling	**AG1**	0.91 (*1.16*) 15%	3.03 (*1.53*) 51%	3.94 (*1.20*) 66%
	AG2	2.00 (*1.62*) 33%	3.92 (*1.44*) 65%	3.96 (*1.20*) 66%
Model story	**AG1**	1.36 (*1.56*) 23%	4.21 (*1.56*) 70%	NA
	AG2	2.08 (*2.04*) 35%	4.50 (*1.10*) 75%	

Note: NA = not available: as model story was not performed at Time 3
* the mean age in each group increased by one year with each test time.

Figure 4 shows the distribution of scores for the comprehension of goals in the telling mode by age group and test time. The analysis revealed a significant development over time in both age groups: from 0.85 points at time1 to 2.24 points at time2 and to 2.76 points at time3 in AG1 ($b = 1.32$, $SE = 0.54$, $t = 2.46$, $p < 0.05$ between time1 and time2; $b = 2.30$, $SE = 0.63$, $t = 3.63$, $p < 0.001$ between time1 and time3); from 1.75 points at time1 to 2.71 points at time3 in AG2 ($b = 2.51$, $SE = 0.80$ $t = 3.14$, $p < 0.01$ between time1 and time3). Given that the older age group scored higher already at time1, it is not surprising that the increase in scores from 1.75 to 2.5 at time2 was not significant. In addition, for AG2 there is an effect of LoE ($b = 0.03$, $SE = 0.01$, $t = 2.40$, $p < 0.05$) which interacts with time at time3 ($b = -0.04$, $SE = 0.02$, $t = -2.51$, $p < 0.05$), suggesting that in this age group the increase in scores between time2 and time3 is less for children with a higher LoE (Table 7).

Figure 5 shows the distribution of scores for the comprehension of goals in the model story mode by age group and test time. The analysis revealed a significant development over time in both age groups ($b = 1.37$, $SE = 0.62$, $t = 2.22$, $p < 0.05$ for AG1; $b = 1.51$, $SE = 0.66$, $t = 2.28$, $p < 0.05$ for AG2) whereas the scores improved from 1.88 points at time1 to 2.79 points at time2 in AG1 and from 2.17 points at time1 to 2.88 points at time2 in AG2. No effects of other variables were found in

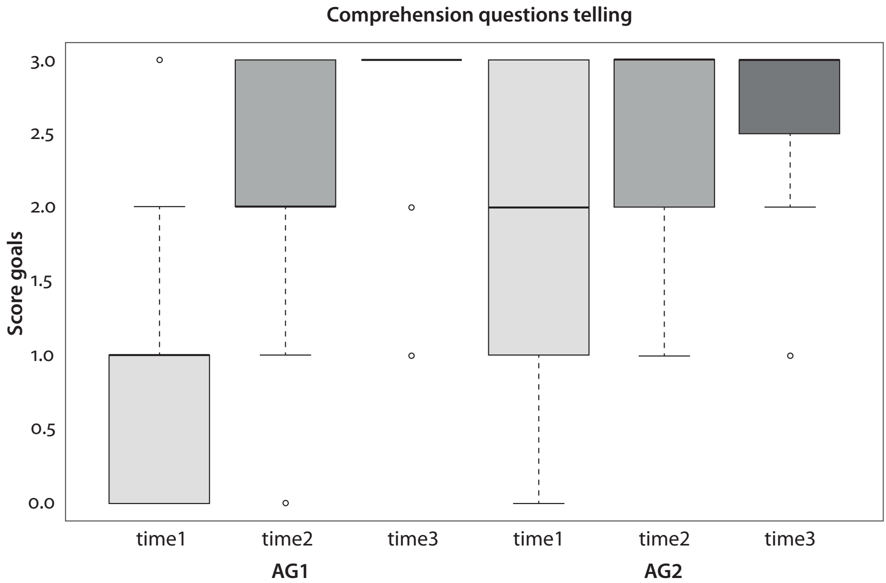

Figure 4. Scores for the comprehension of goals in the telling elicitation mode: Distribution by age group and test time

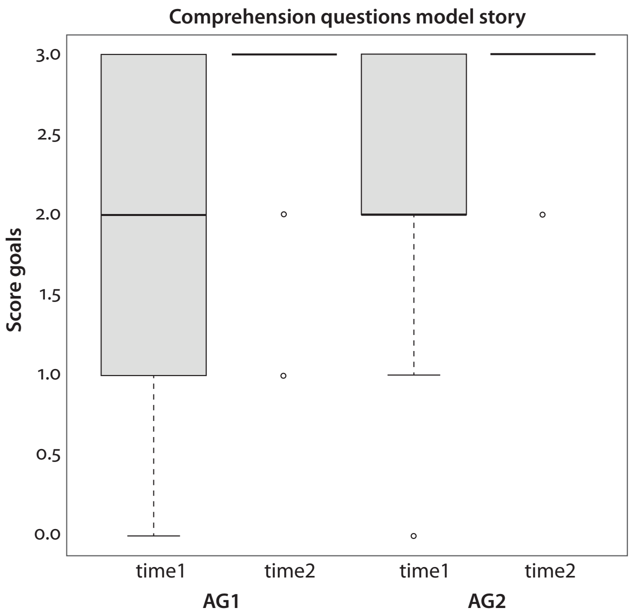

Figure 5. Comprehension of goals in the model story elicitation mode by age group and test time

this analysis (Table 7). Notably, the scores were already higher at time1 in both age groups in comparison to the scores in telling mode.

In the next step developmental trajectories for the comprehension of goals were analysed with regard to the effect of age group and its interaction with time (separately for the telling and model story modes) and with regard to the effect of elicitation mode and its interaction with time (separately for each age group). The developmental trajectories are shown in Figure 6.

With regard to the effect of age group and its interaction with time, the linear regression analysis of scores for the comprehension of goals showed that in the telling mode there are positive main effects of time ($b = 1.37$, $SE = 0.42$, $t = 3.26$, $p < 0.01$ for time2 and $b = 2.77$, $SE = 0.48$, $t = 5.73$, $p < 0.001$ for time3) and age group ($b = 0.84$, $SE = 0.20$, $t = 4.19$, $p < 0.001$), as presented in Table 8. There is also an additional effect of LoE present ($b = 0.02$, $SE = 0.01$, $t = 2.81$, $p < 0.01$). There is a negative interaction between test time and AG2, specifically at time2 ($b = -0.66$, $SE = 0.28$, $t = -2.34$, $p < 0.05$) and at time3 ($b = -0.88$, $SE = 0.28$, $t = -3.10$, $p < 0.01$) as well as a negative interaction between time3 and LoE ($b = -0.03$, $SE = 0.01$, $t = -2.62$, $p < 0.01$). The results suggest, therefore, that the developmental trajectories of the two age groups over the three test times are not similar: the increase is higher in AG1 than in AG2 and children with higher LoE have lower scores at time3. In model story mode there is only a main effect of time ($b = 1.42$, $SE = 0.45$,

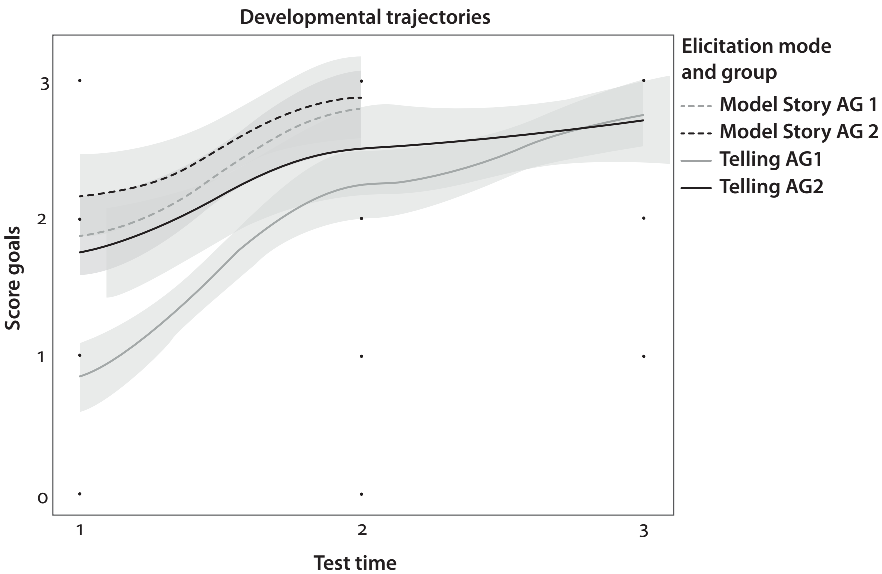

Figure 6. Developmental trajectories of both age groups based on the scores for the comprehension of goals in the telling and model story elicitation modes

Table 7. Linear regression models for developmental trajectories for each elicitation mode and age group separately: Goals

		Age group 1				Age group 2			
		b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Telling	(Intercept)	0.46	0.33	1.42		0.85	0.44	1.95	
	Time2	1.32	0.54	2.46	*	0.85	0.71	1.20	
	Time3	2.30	0.63	3.63	***	2.51	0.80	3.14	**
	LoE	0.02	0.01	1.55		0.03	0.01	2.40	*
	L1t	−0.14	0.26	−0.56		0.12	0.32	0.38	
	Time2:LoE	−0.01	0.01	−0.62		−0.01	0.02	−0.75	
	Time3:LoE	−0.02	0.01	−1.13		−0.04	0.02	−2.51	*
	Time2:L1t	0.57	0.36	1.57		0.17	0.45	0.37	
	Time3:L1t	0.19	0.36	0.53		−0.01	0.45	−0.02	
Model story	(Intercept)	1.61	0.37	4.29	***	1.46	0.41	3.58	***
	Time2	1.37	0.62	2.22	*	1.51	0.66	2.28	*
	LoE	0.01	0.01	0.90		0.02	0.01	1.58	
	L1t	−0.06	0.29	−0.20		0.36	0.30	1.19	
	Time2:LoE	−0.02	0.02	−1.10		−0.02	0.02	−1.20	
	Time2:L1t	0.33	0.42	0.78		−0.42	0.42	−1.00	

Notes: These models include fixed variables of time (time1, time2, time3), LoE, and L1; L1t refers to L1 Turkish as opposed to L1 Russian; significance levels * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

Table 8. Linear regression models with regard to the effect of age group and its interaction with time for each elicitation mode separately: Goals

	Telling				Model story			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
(Intercept)	0.28	0.26	1.10		1.47	0.27	5.36	***
Time2	1.37	0.42	3.26	**	1.42	0.45	3.18	**
Time3	2.77	0.48	5.73	***		NA		
AG2	0.84	0.20	4.19	***	0.23	0.21	1.10	
LoE	0.02	0.01	2.81	**	0.01	0.01	1.62	
L1t	−0.04	0.20	−0.22		0.11	0.21	0.53	
Time2:AG2	−0.66	0.28	−2.34	*	−0.15	0.30	−0.48	
Time3:AG2	−0.88	0.28	−3.10	**		NA		
Time2:LoE	−0.01	0.01	−0.93		−0.02	0.01	−1.50	
Time3:LoE	−0.03	0.01	−2.62	**		NA		
Time2:L1t	0.40	0.28	1.44		0.01	0.30	0.02	
Time3:L1t	0.13	0.28	0.49			NA		

Notes: These models include fixed variables of time (time1, time2, time3), LoE, and L1; L1t refers to L1 Turkish as opposed to L1 Russian; significance levels * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

$t = 3.18$, $p < 0.01$) but no effect of age group. No interaction was found between time and age group, suggesting that the developmental trajectories of the age groups are rather similar: both age groups have the same increase from time1 to time2.

With regard to the effect of elicitation mode and its interaction with time for comprehension questions targeting goals in each age group, the analysis showed that there is a negative effect of the telling elicitation mode in AG1 ($b = -1.03$, $SE = 0.20$, $t = -5.16$, $p < 0.001$) but not in AG2 (Table 9).

Table 9. Linear regression models with regard to the effects of elicitation mode and its interaction with time for each age group separately: Goals

	Age group 1				Age group 2			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
(Intercept)	1.55	0.28	5.61	***	1.36	0.33	4.06	***
Time2	1.10	0.45	2.46	*	1.16	0.54	2.16	*
Mode_Telling	−1.03	0.20	−5.16	***	−0.42	0.23	−1.83	
LoE	0.01	0.01	1.63		0.02	0.01	2.68	**
L1t	−0.10	0.20	−0.50		0.24	0.23	1.03	
Time2:Mode_Telling	0.48	0.28	1.72		0.04	0.32	0.13	
Time2:LoE	−0.01	0.01	−1.18		−0.02	0.01	−1.29	
Time2:L1t	0.45	0.29	1.56		−0.13	0.33	−0.39	

Notes: These models include fixed variables of time (time1, time2, time3), LoE, and L1; L1t refers to L1 Turkish as opposed to L1 Russian; significance levels * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

At the same time, there is a positive effect of LoE for AG2 ($b = 0.02$, $SE = 0.01$, $t = 2.68$, $p < 0.01$). The effect of time is present in both age groups, but it was already analysed separately for each elicitation mode. No interaction between elicitation mode and time was found in either age group, suggesting that both age groups show similar developmental trajectories for both elicitation modes with regard to the comprehension of questions targeting goals.

Figure 7 shows the distribution of scores for the comprehension of internal states in the telling elicitation mode by age group and test time. The analysis of development over time showed that the scores significantly increase from 0.91 points at time1 to 3.03 points at time2 ($b = 1.94$, $SE = 0.97$, $t = 2.00$, $p < 0.05$, Table 10) and to 3.94 points at time3 (not significant though) in AG1, whereas the increase in points from 2.0 at time1 to 3.92 at time2 and 3.96 at time3 in AG2 is not significant for either test time. At the same time, in AG1 an interaction between L1 and time was found ($b = 1.43$, $SE = 0.66$, $t = 2.19$, $p < 0.05$ for time3), indicating that the increase in the scores of Turkish-speaking children in this age group is higher than in those of Russian-speaking children between time2 and time3.

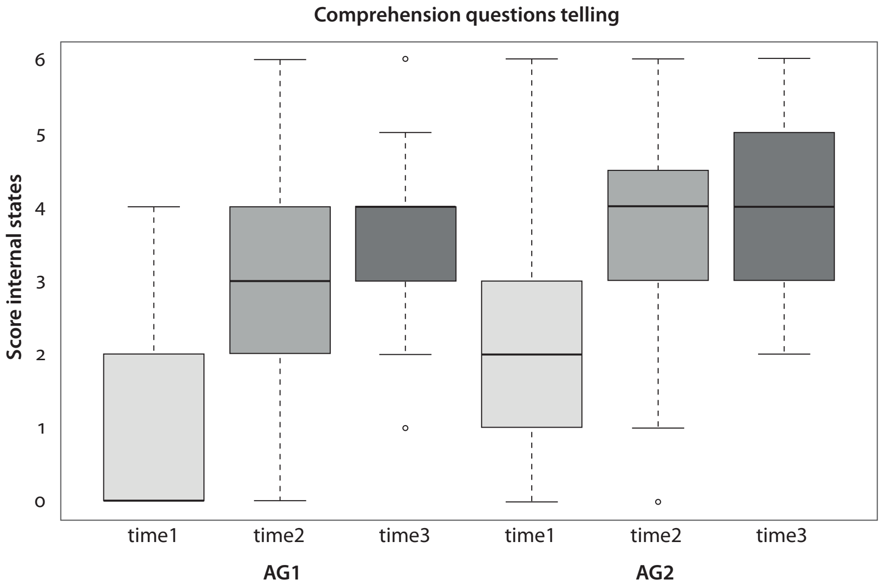

Figure 7. Scores for the comprehension of internal states in the telling elicitation mode: Distribution by age group and test time

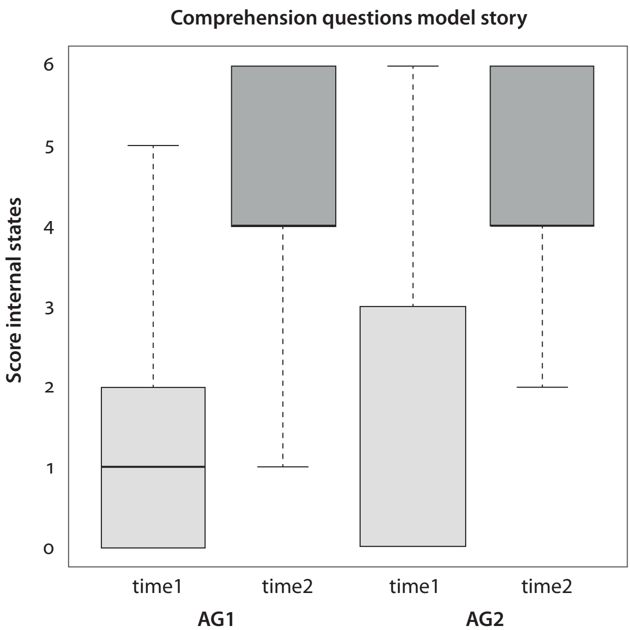

Figure 8. Scores for the comprehension of internal states in the model story elicitation mode: Distribution by age group and test time

Figure 8 presents the distribution of scores for the comprehension of internal states in the model story mode by age group and test time. In comparison to the telling mode, the development over time with regard to internal states is significant for both age groups ($b = 2.36$, $SE = 1.18$, $t = 2.01$, $p < 0.05$ for AG1; $b = 3.15$, $SE = 1.55$, $t = 2.04$, $p < 0.05$ for AG2), with a slight increase in scores: from 1.36 points at time1 to 4.21 points at time2 in AG1 and from 2.08 points at time1 to 4.50 points at time2 in AG2. No effects of other variables were found in this analysis (Table 10).

In order to compare the developmental trajectories for the comprehension of **internal states**, we analysed the effect of age group and its interaction with time (separately for the telling and model story elicitation modes) and the effect of elicitation mode (separately for each age group). The developmental trajectories are shown in Figure 9.

With regard to the effect of age group and its interaction with time in the telling mode, the analysis showed that there are positive main effects of time ($b = 2.03$, $SE = 0.78$, $t = 2.60$, $p < 0.05$ for time2 and $b = 2.74$, $SE = 0.90$, $t = 3.04$, $p < 0.01$ for time3) and age group ($b = 1.11$, $SE = 0.37$, $t = 2.98$, $p < 0.01$, AG2 producing higher scores) and a significant negative interaction between time and age group, specifically with time3 ($b = -1.14$, $SE = 0.53$, $t = -2.18$, $p < 0.05$) as presented in Table 11. The results suggest that the developmental trajectories of the two age groups over the three test times are not similar: the increase is higher in AG1 than in AG2.

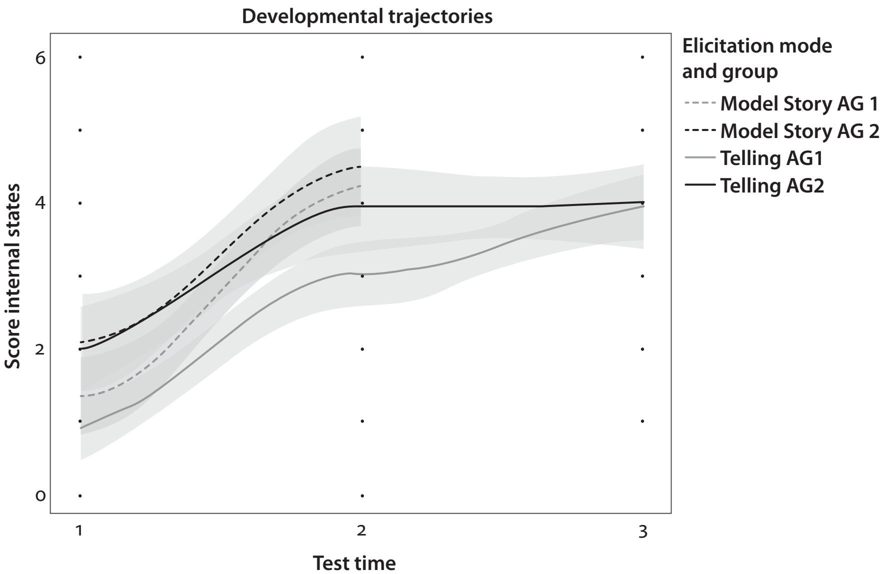

Figure 9. Developmental trajectories of both age groups based on the scores for the comprehension of internal states in the telling and model story elicitation modes

Table 10. Linear regression models for developmental trajectories for each elicitation mode and age group separately: Internal states

		Age group 1				Age group 2			
		b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
Telling	(Intercept)	0.97	0.59	1.66		1.51	0.83	1.82	
	Time2	1.94	0.97	2.00	*	2.19	1.35	1.63	
	Time3	2.27	1.15	1.98		2.55	1.52	1.68	
	LoE	0.01	0.02	0.47		0.02	0.02	0.84	
	L1t	−0.73	0.46	−1.58		−0.12	0.61	−0.19	
	Time2:LoE	−0.01	0.03	−0.30		−0.01	0.03	−0.23	
	Time3:LoE	0.00	0.03	0.04		−0.02	0.03	−0.57	
	Time2:L1t	0.92	0.66	1.41		−0.35	0.86	−0.41	
	Time3:L1t	1.43	0.66	2.19	*	−0.13	0.86	−0.15	
Model story	(Intercept)	1.94	0.71	2.73	**	1.30	0.95	1.36	
	Time2	2.36	1.18	2.01	*	3.15	1.55	2.04	*
	LoE	−0.01	0.02	−0.59		0.02	0.03	0.65	
	L1t	−0.48	0.56	−0.86		0.54	0.70	0.77	
	Time2:LoE	0.01	0.03	0.22		−0.01	0.04	−0.34	
	Time2:L1t	0.88	0.80	1.10		−0.78	0.99	−0.79	

Notes: These models include fixed variables of time (time1, time2, time3), LoE, and L1; L1t refers to L1 Turkish as opposed to L1 Russian; significance levels * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

Table 11. Linear regression models with regard to the effects of age group and its interaction with time for each elicitation mode separately: Internal states

	Telling				Model story			
	b	*SE*	*t*	*p*	*b*	*SE*	*t*	*p*
(Intercept)	0.77	0.48	1.61		1.42	0.57	2.49	*
Time2	2.03	0.78	2.60	*	2.75	0.93	2.98	**
Time3	2.74	0.90	3.04	**		NA		
AG2	1.11	0.37	2.98	**	0.73	0.44	1.66	
LoE	0.01	0.01	0.88		−0.00	0.02	−0.04	
L1t	−0.48	0.37	−1.31		−0.08	0.43	−0.19	
Time2:AG2	−0.23	0.53	−0.45		−0.46	0.62	−0.73	
Time3:AG2	−1.14	0.53	−2.18	*		NA		
Time2:LoE	0.01	0.02	−0.27		0.00	0.02	0.02	
Time3:LoE	0.01	0.02	−0.33			NA		
Time2:L1t	0.38	0.52	0.73		0.19	0.61	0.32	
Time3:L1t	0.79	0.52	1.52			NA		

Notes: These models include fixed variables of time (time1, time2, time3), LoE, and L1; L1t refers to L1 Turkish as opposed to L1 Russian; significance levels * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

For the model story elicitation mode, there is only a main effect of time ($b = 2.75$, $SE = 0.93$, $t = 2.98$, $p < 0.01$) but no effect of age group. No interaction was found between time and age group, suggesting that the developmental trajectories of the age groups with regard to the comprehension of internal states are rather similar: both age groups have the same increase from time1 to time2.

With regard to the effect of elicitation mode and its interaction with time for comprehension questions targeting internal states in each age group, no effects of mode or other variables except for the effect of time ($b = 2.52$, $SE = 0.81$, $t = 3.11$, $p < 0.01$ for AG1 and $b = 2.92$, $SE = 1.09$, $t = 2.69$, $p < 0.01$ for AG2) were found in this analysis (Table 12). The effect of time has already been analysed separately for each elicitation mode. No interaction between elicitation mode and time was found in either age group, suggesting that both age groups show similar developmental trajectories for both elicitation modes with regard to the comprehension of questions targeting internal states.

Table 12. Linear regression models with regard to the effects of elicitation mode and its interaction with time for each age group separately: Internal states

	Age group 1				**Age group 2**			
	b	***SE***	***t***	***p***	***b***	***SE***	***t***	***p***
(Intercept)	1.68	0.50	3.38	***	1.44	0.68	2.12	*
Time2	2.52	0.81	3.11	**	2.92	1.09	2.69	**
Mode_Telling	−0.45	0.36	−1.26		−0.08	0.46	−0.18	
LoE	0.00	0.02	−0.16		0.02	0.02	1.03	
L1t	−0.60	0.37	−1.65		0.21	0.47	0.45	
Time2:Mode_Telling	−0.73	0.51	−1.43		−0.50	0.65	−0.76	
Time2:LoE	0.00	0.02	−0.03		−0.01	0.02	−0.40	
Time2:L1t	0.90	0.52	1.73		−0.56	0.66	−0.85	

Notes: These models include fixed variables of time (time1, time2, time3), LoE, and L1; L1t refers to L1 Turkish as opposed to L1 Russian; significance levels * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$

As the last step of the analysis we compared the comprehension of questions targeting goals and internal states in each elicitation mode and age group. For this comparison the scores were normed to the overall possible score for each category. The results based on the Wilcoxon test confirmed that children of both age groups always showed significantly better scores for goals than for internal states in both elicitation modes at all test times: $V = 211.5$, p-value < 0.05 and $V = 195.5$, $p < 0.001$ at time1, $V = 323.5$, $p < 0.01$ and $V = 130$, $p < 0.01$ at time2, $V = 344$, $p < 0.001$ and $V = 131$, $p < 0.01$ at time3 in AG1 and AG2, respectively (telling elicitation mode); $V = 355.5$, $p < 0.001$ and $V = 268$, $p < 0.001$ at time1 and $V = 190$, $p < 0.001$ and

$V = 120$, $p < 0.001$ at time2 in AG1 and AG2, respectively (model story elicitation mode). These results suggest that goals are easier to comprehend than internal states within this age range, even at age 5;6 and 6;2 (mean age at time3 in AG1 and AG2, respectively) when children do not achieve high scores for questions targeting internal states (only 66% in each age group versus 90% and 92% for goals in AG1 and AG2, respectively).

4.4 A qualitative note on examples from our dataset

The results can be illustrated by several examples. Overall, we found that also within the scores for the same type of episodic structure components (goals or internal states) there are remarkable differences. For example, question D4 targeting the Goal of Episode 2 (*Why does the boy jump upwards?/ Why does the cat climb the tree?*) had the highest accuracy rate among the three questions targeting goals. This can be related to the cognitive transparency of the episode and, possibly also to its placement early on in the story, where it triggers most of the plot. This question was answered correctly by the majority of children (telling: 38, model: 45), e.g., as in Examples (1)–(4):

(1) D4: Why does the cat climb the tree?
Vogel nehmen [to take (the) bird] (correct)
(r100, AG1, time1, telling/Baby Birds, total score 1)

(2) D4: Why does the boy jump upwards?
weil die will den Ballon haben [because she wants to have the balloon] (correct)
(t009, AG1, time1, model/Dog, total score 4)

(3) D4: Why does the fox jump forward?
Kuh essen [to eat (the) cow] (correct)
(t059, AG1, time1, telling/Baby Goats, total score 1)

(4) D4: Why does the fox jump forward?
weil sie den essen möchte [because she would like to eat him] (correct)
(t052, AG2, time1, telling/Baby Goats, total score 5)

The fewest correct answers were given to D7, targeting Goal 3 in the telling elicitation mode. Many children gave correct answers to the first two questions targeting goals (D1 and D4) already at time2 but did not give correct answers to D7, e.g., as in Example (5):

(5) D1: Why was the mother goat in the water?
die will die Baby helfen [she wants to help the baby] (correct)
D4: Why does the fox jump forward?
die will die Schaf fressen [she wants to eat the sheep] (correct)
D7: Why does the bird bite the fox's tail?
die Vogel will die Wolf fressen [the bird wants to eat the wolf] (wrong)
(t008, AG1, time2, telling/Baby Goats, total score 6)

Some of those children who could not answer D7 correctly in the telling mode gave correct answers to D7 in the model story elicitation mode (Cat or Dog stories), thus the difference in understanding of the goal may be partly related to the story content and not to individual differences between children.

Generally, even though children gave correct answers to all the questions targeting goals they still had problems with comprehension of questions targeting internal states. The response accuracies to these questions varied. For example, we found that questions D8 (*Imagine that the dog sees the birds. How does the dog feel?*) and D9 (*Why do you think that the dog feels good?*) were the most difficult of the questions targeting internal states. D8 is the only question starting with *Imagine that…*, relating to a mental state, and the answer cannot be 'read off' the picture. D8 and D9 both address the same type of internal state and are inherently connected.

Some children gave a wrong answer to D8, leading automatically to a score of zero for D9 (which is not asked when the response to D8 is incorrect). This happened even at test time3, i.e. for the oldest children, as illustrated in Example (6):

(6) D8: Imagine that the bird sees the goats. How does the bird feel?
böse [angry] (wrong; expected answer: good, happy, relieved, proud)
(t008, AG1, time3, telling/Baby Goats, total score 7)

Another child answered D8 correctly, however the reason given in the follow-up question D9 was not sufficient, as shown in (7):

(7) D8: Imagine that the dog sees the birds. How does the dog feel?
gut [good] (correct)
D9: Why do you think that the dog feels good?
weil er mag Vögel [because he likes birds] (unclear > zero score)
(r029, AG2, time3, telling/Baby Birds, total score 7)

In this particular case, the child may have misunderstood the goal of the episode concerning the dog grabbing the cat's tail, as the child answers (D7): *weil er will die auch essen* [because he also wants to eat them]. Thus the child thinks that the dog grabbed the cat's tail because he wanted to eat the baby birds as well. However, this mistaken interpretation did not hold for all children who produced wrong answers to D8/D9.

By test time2 and more often by time3, some children give correct answers to both D8 and to D9, as shown in (8) and (9):

(8) D9: Why do you think that the bird is feeling good?
vielleicht sagen die "Danke" [maybe they say 'Thanks'] (correct)
(t058, AG1, time3, telling/Baby Goats, total score 8)

(9) D9: Why do you think that the bird is feeling good?
weil da keins aufgefressen wurde [because nobody was eaten] (correct)
(t022, AG2, time3, telling/Baby Goats, total score 8)

On a general note, some children did not show any narrative comprehension at all in their responses to comprehension questions and thus scored zero points. Interestingly, this only happened at test time1, i.e. with the youngest children, and even when the children had listened to the story before being asked the comprehension questions, as illustrated in Example (10) for the model story mode:

(10) D1: Why does the dog jump forward?
Ich weiß es nicht [I don't know]
D2: How does the dog feel?
Ich weiß es nicht [I don't know]
D3: Why do you think that the dog is feeling angry? (not asked)
-
D4: Why does the boy jump upwards?
Luftballons [balloons]
D5: How does the boy feel?
Ich weiß es nicht [I don't know]
D6: Why do you think that the boy is feeling good? (not asked)
-
D7: Why does the dog grab the sausages?
Ich weiß es nicht [I don't know]
D8: Imagine that the boy sees the dog. How does the boy feel?
Ich weiß es nicht [I don't know]
D9: Why do you think that the boy feels bad? (not asked)
-
(r032, AG1, time1, model/Dog, total score zero)

Such low performance could be explained by a lack of narrative experience or even a lack of general language proficiency in L2 German, in both vocabulary and grammar, so that the child could not answer the comprehension questions at all using full sentences (except for the phrase *Ich weiß es nicht* 'I don't know'). In AG1, children are still very young (in this particular case, the child was aged 3;4 with 21 months LoE to German) and their narrative skills are not yet well developed.

At the same time, several children achieved a total score of 9 already at time2 (i.e. all answers correct). This mainly happened in the model story, i.e. when the children were asked the comprehension questions after they had listened to the story, as illustrated in Example (11):

(11) D1: Why does the cat jump forward?
weil die will den Schmetterling fang [because she wants to catch the butterfly]
D2: How does the cat feel?
schlecht [bad]
D3: Why do you think that the cat is feeling bad?
na weil die sich angesticht hat [well, because she stung herself]
D4: Why does the boy hold the fishing rod in the water?
Na dass er seinen Ball fangt [well, so he catches his ball]
D5: How does the boy feel?
gut [good]
D6: Why do you think that the boy is feeling good?
na weil er die Ball gefangen hat [well, because he caught the ball]
D7: Why does the cat grab the fish?
na weil die will das Fisch essen [well, because she wants to eat the fish]
D8: Imagine that the boy sees the cat. How does the boy feel?
schlecht [bad]
D9: Why do you think that the boy feels bad?
na weil er sieht, dass die Katze das esst [well, because he sees that the cat eats it]
(r082, AG1, time2, model/Cat, total score 9)

In this case, the child clearly shows enough cognitive skills to infer the content from the pictures and from listening to the story. Additionally, this child already understands different types of questions very well. In order to answer correctly, a child should not only understand the pictorial stimuli and the content behind them, but also be able to correctly understand and answer questions, especially wh-questions (for the acquisition of wh-questions in German, see, e.g., Schulz, 2013). Additionally, well elaborated syntactic structures can be seen in the child's answers. The child has acquired the clause-final V-placement of German.[2] So in this example, we see a balanced relationship between the answers to the comprehension questions and the development of grammar. While our data show a diverse picture as far as narrative comprehension is concerned, we were able to note that,

2. In German, V-final position in subordinate clauses is obligatory, and contrasts with the V-second position in main clauses. Its emergence is a sign of advanced language development corresponding to the last milestone in the acquisition of the basic syntactic structure in German (Tracy, 2007).

generally, children who were able to answer the comprehension questions also had a relatively good command of morphology and syntax in their L2, for example, the use of subordinate clauses with *weil* [because], inflected verb forms with correct syntactic placement in main and subordinate clauses, and the use of DPs, albeit with some grammatical gender errors.

5. Discussion and conclusion

The present study has contributed to our understanding of the narrative comprehension skills, measured by macrostructure, in bilingual preschool children. Comprehension of goals and internal states was traced over a period of two years (over three test times) in the L2 German of bilingual Russian- or Turkish-German-speaking children aged 2;10–3;11 (AG1) or 4;0–4;7 (AG2) at the time of the first testing. Comprehension was assessed in two elicitation modes: after presenting visual stimuli and listening to a story told by the examiner – model story mode, and after presenting only visual stimuli and having children tell a story by themselves – telling mode.

For the general development of narrative comprehension, measured as the sum of macrostructure components as well as single scores for goals and internal states, we found a significant development over time in both age groups and both elicitation modes. Interestingly, and taking another perspective – chronological age – we did not find much development after children reached the age 4;10 (when children correctly answered 6 questions out of 9 in the telling mode over three test times). Rather we noted a plateau period after age 4;10. This was true for the both age groups (see Figure 10). The mapping of the results of the two groups on a continuum of chronological age shows similar developmental trajectories for the overlapping age period, starting at 48 months: a steeper development until the age of 58 months where children improve from a mean score of 2.5 (in AG2) and 4 (in AG1) to nearly 6 in both age groups. This developmental overlap as well as comparisons in the introductory chapter of this volume provide additional evidence for age being a crucial factor for the development of narrative comprehension.

Thus, children can be said to reach one of the milestones of narrative comprehension at around age five, when they correctly comprehend goals and internal states which can be inferred from a picture-based (and oral) narrative text. Goals and internal states compose the structure of events and in our stimuli are combined in a *package*, constituting a part of an episode. In order to understand the narrative macrostructure, children need to infer the content from the pictures and to temporally and causally connect single components into a meaningful whole. This process might be connected to the acquisition of sequentially and simultaneously organized

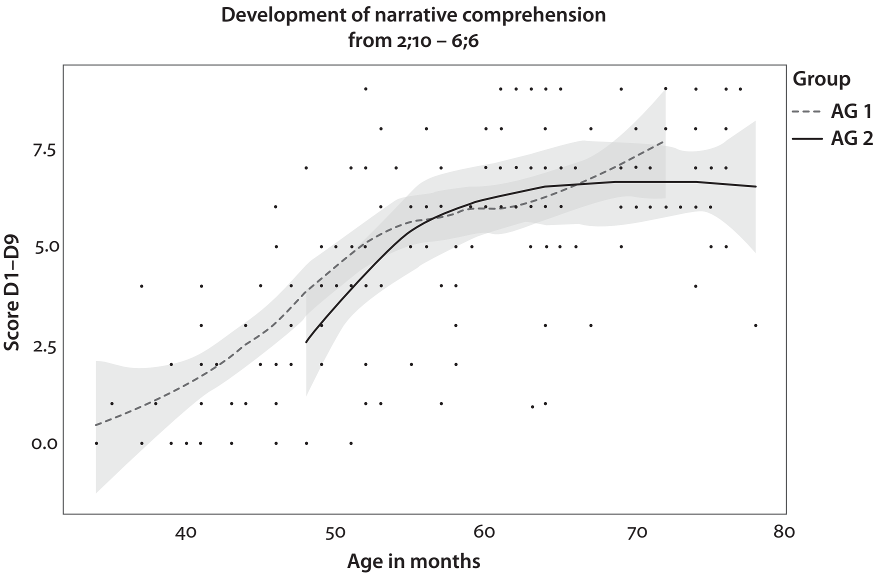

Figure 10. An overlay of the trajectories of narrative comprehension of AG1 and AG2 in the telling elicitation mode

events; if the concept of the organization of discourse is already developed, children can tie the single components of an episode into a complete unit. Such development has been shown to take place before age five by a number of studies; for example, Aksu-Koç and von Stutterheim (1994) showed that children at age five develop an abstract concept of time and can organize events in discourse (cf. Stein, 1988). Similarly to previous studies on narratives reporting considerable improvements in narrative production (cf. Hudson & Shapiro, 1991, etc.), we found a significant spurt in the development of narrative comprehension, however one year earlier, namely at approx. 3;6. This tendency logically follows a path of language acquisition in which comprehension precedes production.

Furthermore, children of both age groups (5- and 6-year-olds) did not reach an absolute ceiling in the comprehension of macrostructure. This was mainly due to failure to provide correct answers to the theory-of-mind question starting with *imagine – Imagine that X sees Y; How does X feel?* – followed by an elaborative question that asks the child to provide a rationale or explanation for this *imagine*-question – *Why do you think that the X feels in that way?* The low accuracies might be due to the complexity of this theory-of-mind question. In this respect, our findings partly corroborate findings from other studies on narrative comprehension assessed with MAIN which demonstrated a significant development over time in comparable age groups, but where even 7-year-olds did not reach the ceiling (e.g., Bohnacker, 2016; Bohnacker & Lindgren, in press; Lindgren, 2018; Lindgren & Bohnacker, this volume; Maviş, Tunçer, & Gagarina, 2016; Bohnacker et al., this volume). Another reason might be that due to the demanding task of imagination, children do not concentrate on the answer itself, but are in a way 'thrown off' the comprehension answer and they concentrate on imagining feelings themselves.

Overall, a steeper development of narrative comprehension has been found between 3;6 and 4;10 children. By age five, children were able to comprehend the majority of the components of an episode they were asked about. These results show that comprehension precedes production because, while at age five children are able to comprehend narrative macrostructure, they still cannot produce it, as indicated in the overview chapter by Berman (2009) on narrative development: "around 5 years <…> [c]hildren will not be able to provide adequate background information or evaluate content to produce a well-motivated, picture-based narrative" (Berman, 2009: 310–311). The majority of children in our study, even with low proficiency in German, were able to comprehend a major part of story organization, 2/3 or more, measured by goals and internal states around age five. This result expands the evidence base for the asymmetry between production and comprehension in language acquisition (Hendriks & Koster, 2010) to the discourse level. Given that children's comprehension of narrative macrostructure can be high even when their macrostructure production is limited (as shown in previous studies on Russian-German bilinguals by Gagarina (2016), narrative comprehension as an assessment tool for bilingual children has good potential – even for children who have insufficient language production. Comprehension of goals and internal states within a story episode presupposes inferencing skills. MAIN is composed of structurally similar episodes, which may still be different concerning 'world-reality', and thus, various types of internal states might be more or less easy to infer. For example, inferring physical pain from a picture of a cat falling into a bush or a dog hitting its head on a tree trunk might be easier, based on children's own experiences of falling or bumping. In contrast, inferencing hunger in a baby bird from a picture of a larger (parent) bird getting a worm (from somewhere) and feeding it to the babies may be less transparent and more demanding as far as cognitive skills and world knowledge are concerned (birds pick and dig up worms and feed them to their young). Thus, episodes may have different 'degrees' of transparency and ease of inferencing, and goals and internal states may be differently comprehended by children of the same age. How exactly different internal states are understood and how this understanding correlates with cognitive skills is an interesting question for future research.

The present study targeted the comprehension of two different components of episodic structure. Protagonists' goals and internal states as reactions to the outcomes of actions were evaluated. The results show that already by time2 in our two age groups, goals are well comprehended in the model story mode and by time3 also in the telling mode. Internal states were comprehended to a much lower degree by children of both age groups. The differences between understanding goals versus internal states were significant in both elicitation modes. The reason for this may be the following. While goals are understood earlier, being the core of macrostructure according to story grammar approaches (Bohnacker, 2016; Trabasso & Nickels, 1992; Trabasso & Rodkin, 1994; among others), internal states, especially as reactions, are generally more complex than other components of story structure as they require not only reflecting on the protagonists' own feelings and thoughts, which, in turn, involves the ability to take the perspective of somebody else (theory of mind). Internal states as reactions require that the child attends to a post-resultative state, i.e. the child reflects on the protagonist's feelings after the action has taken place. While goals also require inferencing, such inferencing is in a way 'more important', since a goal is required to trigger an action, and thus might be more prominent in a child's mind than an internal state resulting from an outcome of an event.

Furthermore, even if children understand the internal state itself, they may have difficulties with reasoning and formulating their answers properly (in such a way that they score a point on MAIN). Thus, first, we can say that comprehension of internal states requires more cognitive and pragmatic skills. Second, answering questions targeting internal states requires more developed lexical and syntactic abilities in their L2, as was underlined by several examples indicating a close relation between grammatical and narrative development. Moreover, we have seen that internal states are not only more difficult, but also that the scores do not improve after time2 in the older age group (between age 5;3 and 6;3). Apparently, children need more time to take another developmental step towards comprehension of internal states such as in the question pair D8/D9 starting with *Imagine that…*, which proved to be the most difficult one, even for older bilingual children (6–7-year-olds) as, e.g., in the study by Bohnacker (2016).

A closer analysis of the interplay of various factors reveals the following picture. The younger group showed a significant effect of time and an interaction between time and L1 (the increase in total scores in the Turkish-speaking children was higher than in the Russian-speaking children) in the telling elicitation mode, as well as an effect of time, but no effects of LoE or L1 or interactions, in the model story elicitation mode. The greater increase in children with L1 Turkish might be due to the fact that their scores were generally lower in the beginning than the scores of children with Russian L1. Basically, in the age range and the period we explored, the cognitive and language development of children seem to enhance the development of narrative comprehension and there are greater effects for children with lower scores at the beginning. It should be noted that the overall number of questions is rather limited and the questions targeting internal states come in pairs, i.e. if the first one is not answered correctly, the second is not asked at all. Thus, a child who fails in answering the first question in such a pair already loses two points while a child who succeeds in answering the first question is more likely to answer the second one as well and score two points more.

The absence of main effects or interactions with LoE with regard to developmental trajectories might be due to the fact that we targeted language comprehension and not language production, and we did not evaluate the correctness of children's grammar. On the other hand, given that the average LoE was 27 and 30 months in AG1 and AG2 respectively, it might have been sufficient to relativize its role at this stage of narrative development. Thus, also a later AoO may not play a considerable role anymore after a certain LoE has been reached and a basic level of language proficiency, needed for comprehension, has been established. It seems that time, and consequently chronological age (which is associated with cognitive maturation), is the most powerful factor and has more impact on changes in the comprehension of macrostructure.

Finally, the present study examined narrative comprehension in two elicitation modes, model story versus telling. Randomization was done in full accordance with the 2012 MAIN guidelines, which recommend the use of Cat and Dog for the model story and Baby Birds and Baby Goats for the telling mode. We found that the total scores are always higher in the model story mode for Cat and Dog (although this difference was not always significant). While previous studies have shown that comprehension of Cat and Dog is higher than of Baby Birds and Baby Goats even when all stories are administered in the telling mode (see Bohnacker & Lindgren, in press, available online via DiVA; Bohnacker, Öztekin & Lindgren, this volume Lindgren, 2018; Lindgren & Bohnacker, this volume; Öztekin, 2019) still the oral support of the model story being told to the child alongside the pictorial stimuli might lead to better comprehension. It is obvious that if a child is proficient in a certain language, listening to a story and viewing the pictures is more profitable than only viewing the pictures. If a child does not understand the whole content of the story through the pictures alone, the story told by the examiner may help to guide attention to pictorial content the child might otherwise not have attended to and explain what the child otherwise would have to infer by her-/himself.

These findings are reminiscent of results from the study of Schneider and Dubé (2005), who compared children's narrative production with regard to different methods of story presentation, oral only, pictures only, and oral-with-pictures, and found that kindergarten children performed best in the oral-with-pictures mode. Also Otwinowska et al. (2018) found that bilingual children (Polish-English) performed better in narrative production and comprehension in the model story mode. However, in the current study, a main effect of elicitation mode was only found in AG1, but not in AG2. Children in AG2 were on average eight months older than children in AG1, and they may have been more attentive to the pictorial stimuli, so that additional oral stimuli played less of a role in the comprehension of macrostructure. Our results corroborate the findings of Maviş et al. (2016), who also compared comprehension in the telling and model story modes and found that Turkish-German bilinguals (three age groups, 2;11–3;11, 4;0–5;11, and 6;0–7;11) had partly higher scores in comprehension questions asked after the model story, especially in the youngest age group. Thus, the increased comprehension after the model story may have several explanations. One could assume that pictorial stimuli displaying the story plot are not enough to allow children of a younger age to fully understand the complexity of the story. When pictures are accompanied by an oral story produced by an examiner, this gives younger children the necessary information about the story plot, goals, and internal states, and facilitates comprehension by verbal support of inferencing. In addition, the complexity of the stories used for the model versus telling elicitation modes is slightly different. The plot in the model stories Cat and Dog is generally easier to interpret than the plot in the Baby Birds and Baby Goats stories.

To conclude, we have found an interesting developmental trajectory of story comprehension over an important period of language acquisition between age 2;11 and 6;6. Children appear to go through a developmental spurt in narrative comprehension and start a plateau period by age 4;10. Moreover, children's comprehension appears to be sensitive to the type of narrative macrostructure component queried, and there are domains where narrative comprehension is not yet fully developed, even for the higher end of the age range at hand. Future studies of narrative comprehension may want to include more detailed, qualitative analyses of story structure components and individual episodes, as they all may be of different complexity.

Results of the analyses of goals and internal states showed different developmental pictures for the two age groups. In AG1, a significant development of goal and internal state comprehension respectively, was found in both modes. In AG2, a significant development in the comprehension of internal states was only found in the model story mode. These results suggest therefore that development in the comprehension of goals and internal states may be modified not only by age (since age overlaps in the two groups), but also by such factors as length of exposure and general level of comprehension.

For future investigations, we plan to compare the narrative production and comprehension in the L1 Russian and Turkish of our bilingual sample. We expect similar results in comprehension in both languages, following the hypothesis that the basic macrostructure is more or less language independent and, once learnt in one language, can be transferred to another (cf. Paradis, Genesee, & Crago, 2011; Pearson, 2002; among others). At the same time, the sample can be adjusted according to different foci in order to investigate the effects of specific factors more precisely and may be compared to samples with other language combinations and target groups.

References

Akinci, M.-A., Jisa, H., & Kern, S. (2001). Influence of L1 Turkish on L2 French narratives. In L. Verhoeven & S. Strömqvist (Eds.), *Narrative development in a multilingual context* (pp. 189–208). Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.23.07aki

Aksu-Koç, A. A., & von Stutterheim, C. (1994). Temporal relations in narrative: Simultaneity. In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: A crosslinguistic developmental study* (pp. 393–456). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bamberg, M. G. W. (1994). Development of linguistic forms: German. In R. Berman & D. Slobin (Eds.), *Relating events in narrative: A crosslinguistic developmental study* (pp. 189–238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bel, A., Ortells, M., & Morgan, G. (2015). Reference control in the narratives of adult sign language learners. *International Journal of Bilingualism*, 19(5), 608–624. https://doi.org/10.1177/1367006914527186

Berman, R. A. (2009). Trends in research on narrative development. In S. Foster-Cohen (Ed.), *Language Acquisition* (pp. 294–318). London: Palgrave Macmillan. https://doi.org/10.1057/9780230240780_13

Berman, R. A., & Slobin, D. I. (Eds.). (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bettge, S., & Oberwöhrmann, S. (2017). *Grundauswertung der Einschulungsdaten in Berlin 2016*. Berlin: Senatsverwaltung für Gesundheit, Pflege und Gleichstellung.

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

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Bohnacker, U., & Lindgren, J. (in press). MAIN story comprehension: What can we expect of a typically developing child? In S. Armon-Lotem & K. Grohmann (Eds.), *LITMUS in Action: Cross-comparison studies across Europe*. Amsterdam: John Benjamins. Preprint retrieved from <http://uu.diva-portal.org/smash/get/diva2:1348899/FULLTEXT01.pdf> (23 June, 2020).

Bohnacker, U., Öztekin, B. & Lindgren, J. (this volume). Bilingual Turkish-Swedish children's understanding of MAIN picture sequences: Individual variation, age, language and task effects. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.61.04boh

De Cat, C. (2008). Experimental evidence for preschoolers' mastery of "topic". In A. Gavarró & M. J. Freitas (Eds.), *Proceedings of GALA 2007* (pp. 155–165). Newcastle upon Tyne: Cambridge Scholars.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools*, 35(2), 155–168. https://doi.org/10.1044/0161-1461(2004/016)

Gagarina, N. (2016). Narratives of Russian-German preschool and primary school bilinguals: Rasskaz and Erzaehlung. *Applied Psycholinguistics*, 37(1), 91–122. https://doi.org/10.1017/S0142716415000430

Gagarina, N., Klassert, A., & Topaj, N. (2010). *Sprachstandstest Russisch für mehrsprachige Kinder / Russian language proficiency test for multilingual children. ZAS Papers in Linguistics,* 54.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, 56.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–276). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-011

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics,* 63.

Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, 120, 1887–1897. https://doi.org/10.1016/j.lingua.2010.02.002

Hickmann, M. (2003). *Children's discourse: Person, space and time across languages*. Cambridge: Cambridge University Press.

Hudson, J. A., & Shapiro, L. R. (1991). From knowing to telling: The development of childen's scripts, stories, and personal narratives. In A. McCabe & C. Peterson (Eds.), *Developing narrative structure* (pp. 89–136). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kapalková, S., Polišenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak–English children: A cross-language comparison. *Applied Psycholinguistics*, 37(1), 145–164. https://doi.org/10.1017/S0142716415000454

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, 37(1), 123–144. https://doi.org/10.1017/S0142716415000442

Leslie, A. M. (1987). Pretense and representation: The origins of 'theory of mind'. *Psychological Review*, 94(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Lindgren, J. (2018). *Developing narrative competence. Swedish, Swedish-German and Swedish-Turkish children aged 4–6* (Studia Linguistica Upsaliensia 19). Uppsala: Acta Universitatis Upsaliensis.

Lindgren, U., & Bohnacker, U. (this volume). Age, language and inferencing: How Swedish-German bilingual preschoolers understand stories. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins.

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish-German bilingual children. *Applied Psycholinguistics*, 37(1), 69–89. https://doi.org/10.1017/S0142716415000429

Meisel, J. M. (2008). Child second language acquisition or successive first language acquisition? In B. Haznedar & E. Gavruseva (Eds.), *Current trends in child second language acquisition. A generative perspective* (pp. 55–80). Amsterdam: John Benjamins. https://doi.org/10.1075/lald.46.04mei

Meisel, J. M. (2011). *First and second language acquisition: Parallels and differences*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511862694

Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 1–25. https://doi.org/10.1080/13670050.2018.1434124

Öztekin, B. (2019). *Typical and atypical language development in Turkish-Swedish bilingual children aged 4–7* (Studia Linguistica Upsaliensia, 25). Uppsala: Acta Universitatis Upsaliensis.

Paradis, J., & Schneider, P. (2008). *Distinguishing bilingual children from monolinguals with SLI: Profile effects on the ENNI*. Poster presented at the Symposium on Research in Child Language Disorders, University of Wisconsin.

Paradis, J. (2019). English second language acquisition from early childhood to adulthood : The role of age, first language, cognitive, and input factors. In M. M. Brown & B. Dailey (Eds.), *BUCLD 43: Proceedings of the 43rd annual Boston University Conference on Language Development* (Vo. 1, pp. 11–26). Somerville, MA: Cascadilla Press.

Paradis, J., Genesee, F., & Crago, M. (2011). *Dual language development & disorders: A handbook on bilingualism & second language learning* (2nd ed.). Baltimore, MD: Paul H. Brookes.

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In K. Oller & R. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853595721-008

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/> (23 June, 2020).

Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian-English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37(1), 49–67. https://doi.org/10.1017/S0142716415000417

Schneider, P., & Dubé, R. V. (2005). Story presentation effects on children's retell content. *American Journal of Speech-Language Pathology*, 14, 52–60. https://doi.org/10.1044/1058-0360(2005/007)

Schulz, P. (2013). Wer versteht wann was? Sprachverstehen im frühen Zweitspracherwerb des Deutschen am Beispiel der w-Fragen. In A. Depperman (Ed.), *Das Deutsch der Migranten* (Jahrbuch des Instituts für Deutsche Sprache 2012) (pp. 313–337). Berlin: De Gruyter. https://doi.org/10.1515/9783110307894.313

Schulz, P., & Grimm, A. (2019). The age factor revisited: Timing in acquisition interacts with age of onset in bilingual acquisition. *Frontiers in Psychology*, 9(JAN), 1–18. https://doi.org/10.3389/fpsyg.2018.02732

Stein, N. L. (1988). The development of children's storytelling skill. In M. B. Franklin & S. S. Barten (Eds.), *Child language: A reader* (pp. 282–291). New York, NY: Oxford University Press.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives* (pp. 53–120). Norwood, NJ: Ablex.

Tellegen, P. J., Laros, J. A., & Petermann, F. (2007). *SON-R 2½-7. Non-verbaler Intelligenztest*. Göttingen: Hogrefe.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, 15, 249–275. https://doi.org/10.1080/01638539209544812

Trabasso, T., & Rodkin, P. C. (1994). Knowledge of goals/plans: A conceptual basis for narrating "Frog where are you?" In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: A cross linguistic developmental study* (pp. 85–106). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tracy, R. (2007). *Wie Kinder Sprachen lernen: Und wie wir sie dabei unterstützen können*. Tübingen: Narr Francke Attempto Verlag.

Tsimpli, I., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, 37(1), 195–216. https://doi.org/10.1017/S0142716415000478

Wagner, R. K., Schatschneider, C., & Phythian-Sence, C. (Eds.). (2009). *Beyond decoding: The behavioral and biological foundations of reading comprehension*. New York, NY: Guilford Press.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, 72, 655–684. https://doi.org/10.1111/1467-8624.00304

Narrative comprehension and its associations with gender and nonverbal cognitive skills in monolingual and bilingual German preschoolers

Carina Marie Wehmeier
Gottfried Wilhelm Leibniz University Hannover

This paper investigates the development of narrative comprehension in monolingual and simultaneously bilingual German preschoolers, in relation to the children's gender and their nonverbal cognitive skills. MAIN (Multilingual Assessment Instrument for Narratives) was used to assess narrative comprehension after telling (Baby Birds) and retelling (Cat) in 199 monolinguals and 66 simultaneous bilinguals. The groups did not differ significantly concerning their nonverbal cognitive skills and German language proficiency. Both the monolingual and the bilingual children were divided into three age groups (4;6–4;11, 5;0–5;5, 5;6–5;11). The narrative comprehension results pointed to significant age effects only for the monolinguals, in both elicitation modes. No gender effects were found for any group. For the nonverbal cognition and narrative comprehension measures, weak correlations obtained for both monolinguals and bilinguals. These results strengthen the idea that narrative comprehension is related to nonverbal cognitive skills of monolingual and bilingual preschoolers.

Keywords: narrative comprehension, nonverbal cognitive skills, gender, monolinguals and bilinguals, preschoolers

1. Narrative skills

Children develop narrative production and comprehension skills during interactions. Peers and adults serve as children's role models in the zone of proximal narrative development when telling stories and provide scaffolding processes when listening to the child's stories, e.g. by asking comprehension questions about relevant story information. Thereby children implicitly derive relevant narrative structures (Nelson, 2010; Nicolopoulou, Brockmeyer, de Sá, & Ilgaz, 2014; Quasthoff

https://doi.org/10.1075/sibil.61.09weh

& Stude, 2018; Vygotsky, 1962). In order for a narrative to be comprehensible, i.e. to be coherent and cohesive, it should be well organized on the two levels: macro- and microstructure (Liles, Duffy, Merritt, & Purcell, 1995; Ringmann, 2013). Microstructure contains language-specific devices that produce cohesion of the sentence elements, e.g. referential and relational elements, while the macrostructure is considered less language-dependent and includes the narrative content and its structural organization (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012, 2015; Gagarina, Klop, Kunnari, Tantele, Välimaa, Bohnacker, & Walters, 2019; Justice, Bowles, Kaderavek, Ukrainetz, Eisenberg, & Gillam, 2006; Pavlenko, 2008; Ringmann, 2014). The macrostructure provides the framework for stories and includes several story structure elements: the *setting*, the time and place of the plot and the introduction of the characters, an *internal state* of a character as an initiating event that leads to a *goal*, which is, in its turn followed by an *attempt* to reach it. The *outcome* follows this action of the character supplemented by an *internal reaction* of the character (Gagarina et al., 2012, 2019; Peterson & McCabe, 1991; Stein & Glenn, 1979; Trabasso & Nickels, 1992). Macrostructural skills can be assessed through narrative production and comprehension tasks that reflect the child's understanding of the story as a whole, of story structure elements, and general knowledge about stories (Westby, 2005). The present paper focuses exclusively on narrative comprehension.

1.1 Narrative comprehension

Narrative comprehension combines knowledge integration (i.e., adaptation of mental representations on the basis of experience) and language. In particular, links are made between the intentions of the characters and the narrative events in the story. Narrative comprehension builds on the skill to generate causal inferences that allow for the completion of non-explicit parts of the story by utilizing world knowledge and understanding of the intentions of characters (Burris & Brown, 2014; Florit, Roch, & Levorato, 2011; Hayward, Schneider, & Gillam, 2009; Kendeou, Bohn-Gettler, White & van den Broek, 2008; Lynch & van den Broek, 2007). Understanding the story structure supports children's narrative comprehension by providing a framework for children to look at a story, listen to a story or recall a story (Hayward et al., 2009). The understanding of the narrative macrostructure can be assessed by comprehension questions that, among other things, capture the child's understanding and awareness of the intentional behavior of the characters (Gagarina et al., 2012). Questions about the story structure help the child to access relevant narrative information (Trabasso, van den Broek, & Liu, 1988) and can enhance the child's understanding of story events and their links by linking the requested information with their knowledge of the story (Anderson, 1994). Different questioning methods (e.g. yes/ no, true/ false, open-ended questions) are used in studies to assess narrative comprehension regarding different story structure elements (Hayward et al., 2009) as well as different narrative methods: comprehension questions after a model story, after retelling a model story and after telling a story (Gagarina et al., 2012, 2019).

There are only a few studies that compare the effects of different narrative methods (retelling vs. telling) on narrative comprehension (e.g. Maviş, Tunçer, & Gagarina, 2016; Otwinowska, Mieszkowska, Białecka-Pikul, Opacki, & Haman, 2018; Roch, Florit, & Levorato, 2016). The performance of children in narrative comprehension tasks, after a picture story has been told to the child (model story) and retold by the child, depends heavily on the capabilities of short-term memory to remember the correct story information. Narrative comprehension may be more difficult after telling a story without listening to a previous model story, as the child may not be aware of all relevant aspects of the story structure due to its level of narrative development (Otwinowska et al., 2018). The results of Maviş et al. (2016), Otwinowska et al. (2018) and Roch et al. (2016) indicate that children have a higher score for narrative comprehension after retelling a picture story than after telling a picture story.

Various studies describe the growth of narrative comprehension in the development of monolingual and bilingual children (Bohnacker, 2016; Hayward et al., 2009; Kapalková, Polisenská, Marková, & Fenton, 2016; Lindgren, 2018; Maviş et al., 2016; Otwinowska et al., 2018; Roch et al., 2016; Westerveld, Gillon, & Boyd, 2012). Burris & Brown (2014) sketch in their review the development of narrative comprehension in children. The authors point to early comprehension skills that develop during infancy as evidenced by Gerson & Woodward (2013) and Sommerville & Woodward (2005). At preschool age, children become more sensitive to goals and relationships between events and begin to integrate their own experiences, leading to an increase in inferences (e.g. Berman & Slobin, 1994; Brown, Lile, & Burns, 2011; Kendeou et al., 2008; Lynch & van den Broek, 2007; Trabasso & Nickels, 1992). The sensitivity of school-aged children to causality and reference to goals increases, and they rely more and more on world knowledge to understand narratives (e.g. Lynch et al., 2008; Lynch & van den Broek, 2007). At about the age of nine, the narrative comprehension skills of a child are similar to those of an adult (e.g. Bohn-Gettler, Rapp, van den Broek, Kendeou & White, 2011; Orrantia, Múñez, & Tarín, 2014; van den Broek, Lynch, Naslund, Ievers-Landis, & Verduin, 2003). In the following, studies on the narrative comprehension of preschoolers are described in more detail.

For example, Westerveld et al. (2012) analyzed the narrative comprehension of 57 typically developing English-speaking children at the age of four and five years longitudinally. The children heard a model story and meanwhile looked at the pictures of the story before answering eight comprehension questions. At the age of four, the children scored significantly lower than at the age of five. Hayward et al. (2009) examined the narrative comprehension of 150 typically developing English-speaking children aged 4 to 6. The authors asked 21 comprehension questions about story structure and the meaning of the story as a whole, after the children retold a story. Overall story comprehension and the accuracy of the answers for both types of comprehension questions were significantly higher for the 6-year-olds than for the 5-year-olds, who again scored significantly higher than the 4-year-olds.

So far, few studies have been conducted on the development of narrative comprehension among monolingual and bilingual children using MAIN used in the present study (e.g. Bohnacker, 2016; Kapalková et al., 2016; Lindgren, 2018; Maviş et al., 2016; Otwinowska et al., 2018; Roch et al., 2016). While most studies focus on bilingual children with a selected language combination (Lindgren, 2018; Otwinowska et al., 2018; Roch et al., 2016), there are very few studies dealing with the narrative development of bilingual children's majority language (e.g. the majority language of the country in which they live) and different additional languages (Boerma, Leseman, Timmermeister, Wijnen, & Blom, 2016; Tsimpli, Peristeri, & Andreou, 2016). For example, Boerma et al. (2016) tested 132 5- and 6-year-old monolingual and bilingual children in Dutch (with and without language impairment) examining, among other things, the impact of bilingualism on their narrative comprehension. The bilingual children had similar narrative comprehension skills when compared to their monolingual peers. Lindgren (2018) analyzed the accuracy of answers to the narrative comprehension questions in 166 monolingual Swedish and bilingual Swedish-Turkish and Swedish-German preschoolers aged 4;0 to 6;11 that were asked after the children had told MAIN stories. She found significant effects of age: regarding the Cat story and the Dog story, six-year-olds scored higher than both other groups without any effect of language group; regarding the Baby Birds story and the Baby Goats story, six-year-olds performed better than five-year-olds, who scored more points than four-year-olds with a significant effect of the language group. Swedish-Turkish children scored lower than monolinguals and Swedish-German children. Lindgren (2018) found significant differences between the stories for monolingual as well as bilingual children: the children's story comprehension scores after telling Cat story and Dog story was higher than after telling Baby Birds story and Baby Goats story. Roch et al. (2016) studied narrative comprehension in 62 Italian-English sequential bilingual preschoolers (M = 5;5 years) and first graders (M = 6;6 years) and found significant effects of age and narrative method. Children scored higher in narrative comprehension when they listened to a story and retold it before answering the questions, as when they were answering comprehension questions after telling a story without listening to it first.

Otwinowska et al. (2018) analyzed the effects of the narrative method (telling vs. retelling) on narrative comprehension among 75 Polish-English and 75 monolingual Polish preschoolers aged three to seven years. The authors reached similar results as Roch et al. (2016): The children answered more comprehension questions correctly after listening to a story and retelling it than after telling a story without listening to it first.

MAIN is a standardized and culturally controlled (not yet norm-referenced) instrument designed with parallel stories of comparable cognitive and linguistic complexity and macrostructure (Gagarina et al., 2012, 2019). To the best of my knowledge, no study has been devoted to the narrative comprehension of MAIN stories in monolingual and simultaneously bilingual (with heterogeneous additional languages) German preschoolers aged 4;6 to 5;11 years. The present paper fills this gap and describes narrative comprehension in German monolinguals and simultaneously bilinguals after telling (Baby Birds story) and retelling a story (Cat story).

1.2 Research on nonverbal cognitive skills and narrative comprehension

Narrative comprehension is based on linguistic skills, phonological, semantic, and syntactic processing as well as the ability to gather and relate information (Szaflarski, Altaye, Rajagopal, Eaton, Meng, Plante, & Holland, 2012). It is believed that narrative comprehension is also associated with cognitive skills such as working memory (Karasinski & Weismer, 2010) and processing capacity and speed (Montgomery, Poluenko, & Marinellie, 2009). Few studies have examined the impact of cognitive skills on children's narrative comprehension (Curenton, 2010; Kim, 2016; Kim, 2015; Montgomery et al., 2009; Murfett, Powell, & Snow, 2008; Schmitter-Edgecombe & Creamer, 2010). These studies may provide an indication of how nonverbal cognitive skills and narrative comprehension are interconnected, which is a focus of this study.

Curenton (2010) studied 72 three- to five-year-old African-American and European-American preschoolers with a low-income family background. Narrative comprehension skills were assessed in English through questions about actions and consciousness (beliefs of the characters) after the experimenter told the story and the child had retold the story. The subscale Language and cognition of Early Screening Inventory-Revised (ESI, Meisels, Marsden, Wiske Stone, & Henderson, 1997) was used to assess cognitive skills. Curenton found a significant effect of question type, with children achieving higher scores answering questions about actions than questions about the consciousness of the characters. In addition, she found a significant age effect. Five-year-olds scored higher than three-year-olds. The narrative comprehension, however, did not predict ESI results, whereas narrative production quality did. In another study, Murfett et al. (2008) examined 78 children between the ages of nine and twelve with a mental disability, age-matched controls and mental age-matched controls. The children should recall a magic show and afterwards answer eight open-ended questions about the story structure. Children with a mental disability proportionally named less story structure elements than children corresponding to age or mental age. Montgomery et al. (2009) analyzed the impact of processing speed on narrative comprehension of 67 children (aged six to eleven years). Auditory-visual response time correlated significantly with the capabilities of narrative comprehension.

These three studies have different setups, samples, and use other tools to evaluate narrative comprehension and cognitive skills than the present study. Nevertheless, the above studies may suggest how nonverbal cognitive skills and narrative comprehension are interconnected. The present study focuses on the relationships between monolingual and simultaneously bilingual German preschoolers' narrative comprehension of MAIN after retelling and telling and nonverbal cognitive skills assessed with the Colored Progressive Matrices (CPM, Raven, Bulheller, & Häcker, 2002).

1.3 Research on gender and narrative comprehension

Studies on gender-specific effects in the various language domains acquisition tend to find that girls up to the age of six perform better on different linguistic tasks than boys (Beltz, Blakemore, & Berenbaum, 2013; Hayiou-Thomas, Dale, & Plomin, 2014; meta-analysis of Hyde & Linn, 1988; Lange, Euler, & Zaretsky, 2016; Ullman, Miranda, & Travers, 2008; Wallentin, 2009). Narratives have also been studied concerning gender differences. However, the results of studies on gender differences in macrostructure production vary (Aldrich & Brooks, 2017; Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004; Gardner-Neblett & Sideris, 2018; Kunnari, Välimaa, & Laukkanen-Nevala, 2016; Maviş et al., 2016). While Maviş et al. (2016) and Kunnari et al. (2016) found no gender differences, Gardner-Neblett & Sideris (2018), Fey et al. (2004) and Nicolopoulou (2008) described gender-specific narrative skills. Different results could be explained by different age and language samples and different narrative assessment tools (Matthews, Biney & Abbot-Smith, 2018). While studies that uncover gender differences tend to assess personal stories which are based on gender differences in experiences (Nicolopoulou, 2008), studies such as those using MAIN elicit fictitious stories with tightly controlled stimulus material and procedure (Maviş et al., 2016; Kunnari et al., 2016) and thus the children might have less leeway to tell divergent stories (Bohnacker & Lindgren, in press).

Even fewer studies have focused on gender differences in narrative comprehension (Aldrich & Brooks, 2017; John, Lui & Tannock, 2003; Maviş et al., 2016).

Aldrich & Brooks (2017) investigated narrative comprehension in 80 English speaking children aged 5;0 to 11;11 years. The children told a wordless picture book before they were asked five comprehension questions, especially about internal states of the characters. There was a significant effect of gender in favor of the girls. Maviş et al. (2016) analyzed the narrative comprehension skills of 36 simultaneously bilingual Turkish-German preschoolers aged 2;11 to 7;11 years with MAIN and compared the results of boys and girls. They found no significant differences between boys and girls in terms of comprehension questions posed both after the model story and after telling a story, but a trend in favor of the girls (slightly higher average scores). John et al. (2003) investigated the narrative comprehension of 61 monolingual English speaking children aged six to eleven years. After retelling a wordless picture book, the children were given five factual (in terms of visualized facts) and five inferential questions (relating to non-explicit parts of the story) regarding the story. They found no gender differences in narrative comprehension, although girls were more accurate in responding to inferential comprehension questions than boys. These diverging outcomes may be explained by different age and languages samples and by different stimulus material and methods for evaluating narrative comprehension (Matthews et al., 2018). The present study wants to compare narrative comprehension after retelling and telling a story using MAIN in monolingual and bilingual girls and boys aged 4;6 to 5;11 years.

2. The present study

The present study is part of an ongoing dissertation project investigating narrative skills, language abilities, auditory perception and processing skills, nonverbal cognitive abilities and the family background in 436 monolingual and bilingual German-speaking children aged 4;0 to 6;11. The children were recruited from 32 kindergartens in the German federal states Lower Saxony, Hesse and Schleswig-Holstein. Preschool teachers were asked to select children for the study based on the kindergarten documentation of child development according to the following criteria: no developmental disabilities in terms of speech, hearing and cognitive skills, and additionally for bilinguals: minimum duration of exposure to German of twelve months. Only children who met the criteria were invited to participate. For each child, written consent of the parents was obtained.

This chapter aims to examine the development of narrative comprehension in 199 monolingual and 66 simultaneously bilingual children aged 4;6 to 5;11 and the relationships between narrative comprehension and nonverbal cognitive skills as well as gender. The following research questions are asked:

1. How does the narrative comprehension of monolingual and simultaneously bilingual children develop between the ages of 4;6 and 5;11?
2. How does elicitation mode – comprehension questions after telling a story vs. after listening and retelling a story – affect the narrative comprehension skills?
3. Are nonverbal cognitive skills associated with the children's narrative comprehension?
4. Is gender in the three age groups associated with narrative comprehension of the children?

Based on previous studies of monolinguals and bilingual preschoolers, I predict significant growth in narrative comprehension with age (Bohnacker, 2016; Hayward et al., 2009; Lindgren, 2018; Maviş et al., 2016; Roch et al., 2016; Westerveld et al., 2012). For the second research question, I foresee significant differences in narrative comprehension after telling a story and after listening and retelling a story, as some studies have found that after retelling a story, children answered more comprehension questions correctly than after telling a story without first listening to it (Maviş et al., 2016; Otwinowska et al., 2018; Roch et al., 2016).

The impact of nonverbal cognitive skills assessed with the Coloured Progressive Matrices (CPM, Raven et al., 2002) is difficult to predict. Basic cognitive skills, especially working memory, are important for various linguistic skills (Melzer, Rißling, & Petermann, 2016). Although Curenton (2010), Montgomery et al. (2009) and Murfett et al. (2008) found that cognitive skills affect narrative comprehension, these studies have different setups and use different instruments and samplings than this study. Crystalline intelligence,[1] for example, was evaluated in the study of Curenton (2010), taking into account working memory and the extent and depth of (mostly) linguistically represented knowledge (Renner & Mickley, 2015). The CPM, by contrast, contains only one task type, representing only fluid intelligence. Fluid intelligence includes, for example, skills such as inductive thinking and sequential reasoning, as well as the ability to solve unfamiliar problems through targeted reasoning and logical thinking (Mickley & Renner, 2010). There may be correlations between nonverbal cognitive skills and narrative comprehension regarding items that target these abilities, such as explanation questions.

It is debatable whether girls have an advantage in linguistic development over boys (Hamilton, Plunkett, & Schafer, 2000; Maital, Dromi, Sagi, & Bornstein, 2000; Szagun, 2006). Based on previous research, (potential) gender differences

1. Crystalline intelligence describes the range of the acquired, culture-related knowledge, e.g. vocabulary, general language-based declarative and procedural knowledge, language comprehension and knowledge of culturally relevant knowledge areas and the ability to apply that knowledge (Mickley & Renner, 2010).

for narrative comprehension are difficult to predict: while Aldrich & Brooks (2017) found significant gender-specific differences when inquiring internal states of characters, John et al. (2003) and Maviş et al. (2016) found no significant effects of gender in narrative comprehension. Because Maviş et al. (2016) also used MAIN, I do not foresee any gender differences in narrative comprehension either.

3. Methods

3.1 Participants

The study was conducted in the children's kindergartens in fall 2016 and fall 2017, after written parental consent had been obtained. The parents agreed to complete a paper and pencil questionnaire on their child's and family background. The questionnaire was developed as part of this dissertation project and administered in German. The experimenters, who passed on the questionnaire to the parents in the kindergartens, offered help filling in the questionnaire.

Table 1. Mean age in months (SD: standard deviation) and gender (m: male, f: female) of the monolingual and simultaneously bilingual children's three age groups

Sample	**Monolingual children**			**Simultaneously bilingual children**			***all***
	m	***f***	***all***	***m***	***f***	***all***	
4;6–4;11 years	23	31	54	8	9	17	71
Mean age	57.2	56.7	56.9	57.4	56.3	56.8	56.9
(SD)	(1.9)	(1.6)	(1.7)	(2.2)	(1.4)	(1.8)	(1.8)
5;0–5;5 years	36	35	71	16	10	26	97
Mean age	62.8	63.2	63.0	63.2	62.3	62.9	63.0
(SD)	(1.7)	(1.9)	(1.8)	(1.7)	(1.8)	(1.7)	(1.8)
5;6–5;11 years	30	44	74	7	16	23	97
Mean age	68.8	68.8	68.8	67.7	69.3	68.8	68.8
(SD)	(1.5)	(1.7)	(1.6)	(1.6)	(1.9)	(2.0)	(1.7)
all	89	110	199	31	35	66	265

The experimenters administering MAIN and CPM were trained master students of rehabilitation pedagogy and native speakers of German. Children and experimenters met in the child's kindergarten group, where the experimenters spent time building a trusting relationship with each child. When the child verbally consented to participate in the study, assessment took place individually in a separate room. The assessment was stopped when the child showed signs of fatigue.

Data from 199 monolingual and 66 simultaneously bilingual German boys and girls aged 4;6 to 5;11 years (*M* = 63 months, *SD* = 5 months) were included in this analysis (Table 1). Some children (monolinguals *N* = 45, bilinguals *N* = 13) did not complete all parts of the assessment because of illness or fatigue, which is why these data were excluded. The preschoolers were divided into three age groups (AG1: 4;6–4;11 years, AG2: 5;0–5;5 years, AG3: 5;6–5;11 years). On average, the monolinguals and simultaneous bilinguals of AG1 are 56.9 months (*SD* = 1.8 months), AG2 63.0 months (*SD* = 1.8 months) and AG3 68.8 months (*SD* = 1.7 months) old. The average age difference between AG1 and AG2 is 6.1 months, while that between AG2 and AG3 is 5.8 months (Table 1).

The parental questionnaire provided information about the linguistic background. The parents of all bilingual children stated that their child had regular contact with the German language from birth (age in months thus corresponds to length of exposure to German). The languages that the bilingual children spoke in addition to German were heterogeneous: 23 different language combinations are part of this sample. The most frequent languages were Russian (*N* = 16), English (*N* = 14) and Polish (*N* = 9). In total, twelve children were raised trilingual from birth. 50% of the simultaneously bilingual children's parents stated that they spoke equal amounts of German and of the other family language at home, while the other half of the parents stated that they mostly spoke the other family language (i.e. not German) at home. On average, the bilinguals of AG1 attended the German kindergarten for 27 months (*SD* = 9 months), AG2 for 38 months (*SD* = 14 months) and AG3 for 39 months (*SD* = 10 months).

The parental questionnaire also provided information on language development of the children. 85% of the monolinguals and 90% of the bilinguals were around the age of one when they spoke their first German word. Around the age of two, 93% of the monolinguals and 96% of the bilinguals spoke their first sentences (two-word phrases) and had an active German vocabulary of about 50 words. At the time of data collection, parents rated their child's receptive language (monolinguals: 94%, bilinguals: 90%) and productive language (monolinguals: 96%, bilinguals: 91%) in German as "good". The linguistic perception and processing skills of the participants were assessed by means of the Heidelberger auditives Screening in der Einschulungsuntersuchung (HASE, Schöler, & Brunner, 2008). The test measures the auditory perception and processing (repetition of sentences, sequences of numbers and non-words). The monolingual and simultaneously bilingual children's age groups did not differ statistically (Mann-Whitney-U-tests) in terms of the number of correctly repeated sentences, number sequences and non-words (Table 2). The German language skills of the bilinguals are therefore not significantly lower than those of their monolingual peers, at least measured by the HASE.

Table 2. Means (SD: standard deviation) of auditory perception and processing subtests (HASE, Schöler & Brunner, 2008) of the monolingual (MO) and simultaneously bilingual (BI) children's three age groups

HASE	4;6–4;11 years		5;0–5;5 years		5;6–5;11 years	
T-scores:* *M = 50, SD = 10*	**MO** *N = 67*	**BI** *N = 20*	**MO** *N = 80*	**BI** *N = 28*	**MO** *N = 82*	**BI** *N = 28*
Sentence repetition (SD)	50.1 (9.6)	48.9 (10.3)	48.3 (8.7)	44.7 (7.3)	46.3 (8.9)	44.3 (9.7)
Number sequence repetition (SD)	50.7 (11.6)	49.7 (12.7)	52.8 (10.5)	51.5 (10.5)	51.4 (10.3)	47.7 (8.4)
Nonword repetition (SD)	47.1 (10.2)	48.2 (10.2)	46.9 (9.5)	48.2 (8.4)	45.6 (9.7)	46.7 (12.5)

* Based on the empirical results of norm samples, the raw values of tests can be converted into standardized values such as T-scores ($M = 50$, $SD = 10$). In this way, the test result of a person can be estimated in comparison to the norm sample (Döring & Bortz, 2016).

The parental questionnaire also provided information about parents' educational background. The educational background of both parents of a child was scored on a 5-point scale: (1) no graduation certificate, (2) lower secondary school leaving certificate, (3) secondary school leaving certificate, (4) advanced technical college entrance qualification/ higher school leaving certificate, (5) Bachelor/ Master/ other university degree. 51% of the mothers and 48% of the fathers had an advanced technical college entrance qualification/ higher school leaving certificate or a university degree. This indicates that parental education levels in the sample were higher than for the German population as a whole (Bildungsbericht, 2018), which should be taken into account when interpreting the results of this study. The monolingual and the simultaneously bilingual groups did not differ in the educational background of the mothers ($U(59, 171) = 4541$, $p = .23$) and fathers ($U(59, 167) = 4453$, $p = .25$).

3.2 MAIN

Two picture sequences from the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2012, 2015, 2019) of the Language Impairment Testing in Multilingual Settings test battery (LITMUS, Armon-Lotem et al., 2015) were used to evaluate the children's narrative comprehension skills. Each story contains six pictures that illustrate the three episodes of the story. Each episode consists of an internal state of the character as an initiating event leading to a goal. An action attempt of the character is followed by the outcome and complemented by an internal state of the character in reaction to it. The three episodes of Cat are about a cat that wants to catch a butterfly, a boy who wants to get his ball out of the water and the cat that wants to eat the boy's fish. The three episodes of Baby Birds are about a mother bird that wants to feed its babies, a cat that wants to catch one of the mother bird's babies and a dog that wants to protect the birds from the cat (Gagarina et al., 2012, 2019). Narrative comprehension was assessed with ten standardized comprehension questions for each MAIN story. Three comprehension question are related to the three goals (G) of the main characters (one in each of the three episodes). Three questions triggered internal states (IS) of the characters associated with either the initiating event or the reaction element and three questions triggered the child's explanations (E) for these internal states of the characters. The last question examined theory of mind (ToM) to determine if the child can grasp the meaning of the entire story (Table 3, Gagarina et al., 2012, 2019).

Within the same test session, both stories were administered. The narrative comprehension was first evaluated in the Cat story after the child listened to the story and retold it while looking at the pictures. After that, the child told the Baby Birds story from the pictures without having heard the story before, and then the comprehension questions were asked.

The assessment of MAIN was audio recorded, and each story and the answers to the following comprehension questions were transcribed by a trained master students of rehabilitation pedagogy or the author, using Codes for Human Analysis of Transcripts conventions (CHAT, MacWhinney, 2000). Fourteen percent of the audio files ($N = 38$) were additionally transcribed by another trained transcriber and interrater reliability was determined via Cohen's Kappa. The reliability value was 90.1%. The answers to the ten comprehension questions for each story were analyzed by trained master students of rehabilitation pedagogy or the author according to MAIN procedures and classified as correct or wrong. Each correctly answered comprehension question was therefore scored one point, which corresponds to a maximum score of ten points. Fourteen percent of the children's answers to comprehension questions ($N = 38$) were analyzed twice and the interrater reliability (Cohen's Kappa) was 89.3%.

Since three comprehension questions focused on the three goals, three questions elicited internal states and three questions elicited the child's explanations for these internal states, three subscores were analyzed separately: goal comprehension (G), internal state comprehension (IS), and explanations (E), each with a maximum of three points (Table 3).

Table 3. Comprehension questions and correct example answers regarding the Cat and Baby Birds stories (MAIN, Gagarina et al., 2012)

		Cat story		**Baby Birds story**	
		Questions	**Example answers**	**Questions**	**Example answers**
Episode 1	G	Why does the cat jump/ leap forward?	It wants to catch the butterfly	Why does the mother bird fly away?	It wants to get food for its babies
	IS	How does the cat feel?	Angry	How do the baby birds feel?	Hungry
	E	Why do you think that the cat is feeling angry?	It couldn't catch the butterfly	Why do you think that the baby birds are feeling hungry?	They are asking for food
Episode 2	G	Why does the boy hold the fishing rod in the water?	He wants to get his ball back	Why is the cat climbing the tree?	It wants to eat a baby bird
	IS	How does the boy feel?	Happy	How does the cat feel?	Still hungry
	E	Why do you think that the boy is feeling happy?	He got the ball back	Why do you think that the cat is feeling hungry?	It did not get the baby birds
Episode 3	G	Why does the cat grab the fish?	It wants to eat the fish	Why does the dog grab the cat's tail?	It wants rescue the baby bird
	IS	Imagine that the boy sees the cat. How does the boy feel?	Mad	Imagine that the dog sees the birds. How does the dog feel?	Proud
	E	Why do you think that the boy feels mad?	The cat ate his fish	Why do you think that the dog feels relieved?	It saved the birds
	ToM	Will the boy be friends with the cat? Why?	No, because the cat ate his fish	Who does the mother bird like best, the cat or the dog? Why?	The dog, because it saved the baby bird

3.3 CPM

Nonverbal cognitive abilities were assessed using Colored Progressive Matrices (CPM, Raven et al., 2002), which can be administered to children aged three to eleven years. The CPM was administered as the last task (after MAIN). The CPM contains three sets of twelve pattern tasks of increasing difficulty. The order of the sets and the entire process of assessment are standardized. In total, the children were asked 36 times to choose the right one out of six patterns to fill a gap in a large colored pattern image. The nonverbal cognitive abilities of CPM were scored by summing up the raw scores of the correct answers of the three sets. This sum of raw scores was converted into percentile ranks (Raven et al., 2002).

4. Results

4.1 Narrative comprehension

First, results regarding the first research question are presented. The mean values (Figure 1) and the results of the ANOVAs of the comparisons of the narrative comprehension skills (maximum of ten points) of the three age groups (AG1: 4;6–4;11 years, AG2: 5;0–5;5 years, AG3: 5;6–5;11 years) are reported.

The analyses for narrative comprehension revealed statistically significant main effects of the age group after retelling ($F(2, 198) = 8.742$, $p < .001$, $\eta^2 = 0.133$) and after telling ($F(2, 198) = 17.160$, $p < .001$, $\eta^2 = 0.156$) for monolingual children. For both tasks, post-hoc analyses showed that the youngest children's results (after retelling: $M = 5.7$, $SD = 1.6$, after telling: $M = 3.9$, $SD = 1.4$) were significantly

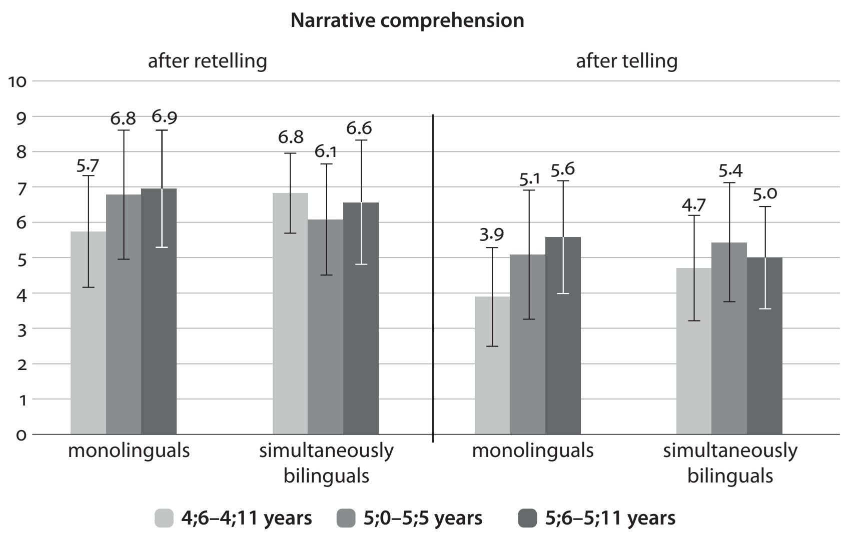

Figure 1. Narrative comprehension of the monolingual and simultaneously bilingual children

Note. Mean values (max. 10 points) and standard deviations (error bars) of the three monolingual and simultaneously bilingual age groups (4;6–4;11, 5;0–5;5, 5;6–5;11 years) after retelling Cat and after telling Baby Birds.

lower than those of the other two groups (AG2: after retelling: $M = 6.8$, $SD = 1.8$, after telling: $M = 5.1$, $SD = 1.8$, AG3: after retelling: $M = 7.0$, $SD = 1.7$, after telling: $M = 5.6$, $SD = 1.6$).

For the three simultaneously bilingual age groups, there were no significant age effects on comprehension questions (after retelling: $F(2, 65) = 1.321$, $p = .274$, after telling: $F(2, 65) = 1.142$, $p = .326$). There were no significant differences between the number of correctly answered questions of the youngest group (after retelling: $M = 6.8$, $SD = 1.1$, after telling: $M = 4.7$, $SD = 1.75$), the middle group (after retelling: $M = 6.1$, $SD = 1.6$, after telling: $M = 5.4$, $SD = 1.7$) and the oldest children (after retelling: $M = 6.6$, $SD = 1.8$, after telling: $M = 5.0$, $SD = 1.6$).

Detailed analyses of the three comprehension subscores (goal questions, internal state questions, explanation questions, with a maximum of three points each) were carried out (Table 4). Some significant age effects were found in monolingual children after retelling (goals: $F(2, 198) = 3.356$, $p = .037$, $\eta^2 = .062$, internal states: $F(2, 198) = 6.452$, $p = .002$, $\eta^2 = .092$, explanations: $F(2, 198) = 3.581$, $p = .030$, $\eta^2 = .048$) and after telling a story (goals: $F(2, 198) = 6.289$, $p = .002$, $\eta^2 = .075$, explanations: $F(2, 198) = 12.374$, $p < .001$, $\eta^2 = .115$). After retelling a story, the youngest monolingual children gave significantly less accurate response to goal questions ($M = 1.9$, $SD = 0.9$) and explanation questions ($M = 1.4$, $SD = 0.5$) than monolingual AG2 (goals: $M = 2.3$, $SD = 0.9$, explanations: $M = 1.6$, $SD = 0.6$).

Table 4. Means (*SD*: standard deviation) of the children's total number of correctly answered comprehension questions of the three subscores (maximum of 3 points each) goals (*G*), internal states (*IS*) and explanations (*E*) after retelling Cat and telling Baby Birds

Narrative comprehension		After retelling			After telling		
		4;6–4;11	5;0–5;5	5;6–5;11	4;6–4;11	5;0–5;5	5;6–5;11
Monolinguals *N = 199*	G	1.9[a] (0.9)	2.3[b] (0.9)	2.1[a,b] (0.9)	1.9[c] (0.9)	2.2[c,d] (0.8)	2.4[d] (0.8)
	IS	2.2[a] (0.9)	2.4[a,b] (0.8)	2.7[b] (0.6)	1.4 (0.9)	1.6 (0.9)	1.8 (0.8)
	E	1.4[a] (0.5)	1.6[b] (0.6)	1.6[a,b] (0.6)	0.0[c] (0.2)	0.5[d] (0.8)	0.5[d] (0.7)
Simultaneously bilinguals *N = 66*	G	2.2 (0.8)	1.8 (0.8)	1.8 (0.9)	2.1 (0.7)	2.2 (0.7)	2.2 (0.7)
	IS	2.8 (0.6)	2.5 (0.7)	2.7 (0.6)	1.8 (0.9)	1.8 (0.7)	1.7 (0.8)
	E	1.5 (0.6)	1.4 (0.7)	1.7 (0.7)	0.3 (0.5)	0.5 (0.6)	0.4 (0.7)

Note. Within each elicitation mode, different indices on two mean scores indicate that there is a significant difference between these scores. Identical indices indicate no significant difference.

The score of the youngest monolingual children in terms of internal state questions (*M* = 2.2, *SD* = 0.9) was significantly lower than that of the oldest monolingual children (*M* = 2.7, *SD* = 0.6).

After telling a story, the youngest monolingual children scored significantly lower in answering goal questions (*M* = 1.9, *SD* = 0.9) than the oldest monolingual children (*M* = 2.4, *SD* = 0.8). Their answers to explanation questions (*M* = 0.0, *SD* = 0.2) were significantly lower than for both other monolingual age groups (AG2: *M* = 0.5, *SD* = 0.8, AG3: *M* = 0.5, *SD* = 0.7). There was no significant age effect with respect to internal state questions (AG1: *M* = 1.4, *SD* = 0.9, AG2: *M* = 1.7, *SD* = 0.9, AG3: *M* = 1.8, *SD* = 0.8).

There were no significant age effects for simultaneously bilingual children when they answered comprehension questions after retelling Cat in relation to the three comprehension subscores: goal questions (AG1: *M* = 2.2, *SD* = 0.8, AG2: *M* = 1.9, *SD* = 0.8, AG3: *M* = 1.8, *SD* = 0.9), internal state questions (AG1: *M* = 2.8, *SD* = 0.6, AG2: *M* = 2.5, *SD* = 0.7, AG3: *M* = 2.7, *SD* = 0.6) and explanation questions (AG1: *M* = 1.5, *SD* = 0.6, AG2: *M* = 1.4, *SD* = 0.7, AG3: *M* = 1.7, *SD* = 0.7). Similarly, there were no significant age effects for simultaneously bilingual children answering comprehension questions after telling Baby Birds in relation to the three comprehension subscores: goal questions (AG1: *M* = 2.1, *SD* = 0.7, AG2: *M* = 2.2, *SD* = 0.7, AG3: *M* = 2.2, *SD* = 0.7), internal state questions (AG1: *M* = 1.8, *SD* = 0.9, AG2: *M* = 1.9, *SD* = 0.7, AG3: *M* = 1.7, *SD* = 0.8) and explanation questions (AG1: *M* = 0.3, *SD* = 0.5, AG2: *M* = 0.5, *SD* = 0.6, AG3: *M* = 0.4, *SD* = 0.7).

Detailed analyses via repeated-measures ANOVAs were carried out to compare the children's skills with regard to the three comprehension subscores (goal questions, internal state questions, explanation questions) with a maximum of three points each (Table 3). Looking at the mean values of the subscores after retelling the story of all monolinguals and simultaneous bilinguals, all children answered more internal state questions correctly than goal questions, which in turn were more often correct than explanations (explanations < goals < internal states; monolinguals: $F(2, 396) = 92.767$, $p < .001$, $\eta^2 = 0.319$, bilinguals: $F(2, 130) = 42.250$, $p < .001$, $\eta^2 = 0.394$). In contrast to that the comprehension pattern after telling a story was that the children answered more goal questions than internal state questions correctly, which were more often correct than explanation questions (explanations < internal states < goals, monolinguals: $F(2, 396) = 320.857$, $p < .001$, $\eta^2 = 0.618$, bilinguals: $F(2, 130) = 139.204$, $p < .001$, $\eta^2 = 0.682$).

Let us now turn to the results of the second research question on the impact of the elicitation mode – comprehension questions after telling vs. after listening and retelling – on comprehension skills. Overall, repeated-measures ANOVAs showed significant effects on the method of assessing comprehension. These effects were strong for monolinguals ($F(1, 198) = 133.981$, $p < .001$, $\eta^2 = 0.404$) and for bilinguals ($F(1, 65) = 30.356$, $p < .001$, $\eta^2 = 0.318$). All children answered more comprehension questions (CQ) correctly after listening to Cat and retelling it than after telling Baby Birds without listening to it first.

The detailed analyses in relation to the elicitation mode focused on the three comprehension subscores (goal questions (G), internal state questions (IS), explanations (E)), which followed the above-mentioned pattern (narrative comprehension after retelling > after telling) with respect to internal state questions (monolinguals: $F(1, 198) = 133.154$, $p < .001$, $\eta^2 = 0.411$, bilinguals: $F(1, 65) = 66.441$, $p < .001$, $\eta^2 = 0.505$) and explanations (monolinguals: $F(1, 198) = 421.285$, $p < .001$, $\eta^2 = 0.680$, bilinguals: $F(1, 65) = 107.586$, $p < .001$, $\eta^2 = 0.623$). However, there were no assessment specific differences in relation to goal questions (monolinguals: $F(1, 198) = 0.280$, $p = .597$, bilinguals: $F(1, 65) = 3.385$, $p = .070$).

An effect of the method of assessing narrative comprehension (after retelling vs. after telling) can only be partially modeled for the age groups: for the total comprehension score (CQ), there are significant differences for all monolingual age groups and at the same time bilinguals AG1 and AG3. The children's comprehension scores were higher after retelling than after telling a story. There are significant differences for the subscores internal state questions and explanation questions for all monolingual and simultaneously bilingual age groups. There is a significant effect of narrative method for the subscore goal questions of for monolinguals AG3, but not for all other monolingual and simultaneously bilingual age groups.

4.2 Impact of nonverbal cognitive skills on narrative comprehension

Now, the third research question about the relationship between nonverbal cognitive abilities and narrative understanding is discussed. The results of the children of the Colored Progressive Matrices test (CPM) are similar for simultaneously bilinguals (raw score: $M = 18.6$, $SD = 4.2$, mi$n = 11$, max = 30, percentile rank: $M = 59.8$, $SD = 28.2$, min = 3, max = 100) and monolinguals (raw score: $M = 17.8$, $SD = 4.3$, min = 7, max = 30, percentile rank: $M = 53.5$, $SD = 28.6$, min = 1, max = 100). Mann-Whitney-U-tests showed no significant differences between the two groups (raw scores: $U(66, 199) = 5925$, $p = .23$, percentile ranks: $U(66, 199) = 5704$, $p = .11$).

The following presents the associations (Pearson's bivariate correlations) between nonverbal cognitive skills (CPM) and narrative comprehension scores (Table 5). Although some correlations are statistically significant, the correlation coefficients are rather low. If several separate univariate analyses are performed, the significance level should be adjusted, e.g. with the Bonferroni correction (Döring & Bortz, 2016). The correlations still significant after the Bonferroni correction are

Table 5. Correlations between the monolingual and simultaneously bilingual children's narrative comprehension (after retelling Cat and telling Baby Birds, total comprehension score (CQ), subscores: goals (G), internal states (IS), explanations (E)) and their nonverbal cognitive skills (raw score CPM)

Narrative comprehension		After retelling				After telling			
		CQ	G	IS	E	CQ	G	IS	E
CPM	Monolinguals	**.21****	.15*	**.18****	.13	**.26****	.18*	.12	**.20****
	Simult. bilinguals	–.07	–.04	.02	–.11	.28*	.10	.14	**.32****

Note
** Correlation is statistically significant (.01)
* Correlation is statistically significant (.05).
Bold-face indicates that the correlation remains statistically significant after Bonferroni correction.

bolded in Table 5. After the Bonferroni correction, there are significant correlations for the monolinguals' narrative comprehension and their nonverbal cognitive skills (after retelling: CQ: r = .21**, IST: r = .18**, after telling: CQ: r = .26**, E: r = .20**), whereas for the bilinguals only the narrative comprehension subscore explanations after telling a story correlated significantly with nonverbal cognitive skills (after telling: E: r = .32**).

4.3 Impact of gender on narrative comprehension

Finally, the fourth research question about the impact of gender on narrative comprehension skills is addressed. Gender differences in narrative comprehension were analyzed in monolingual and simultaneously bilingual children for both elicitation modes and for every age group. All age groups were compared overall via ANOVAs and individually (via non-parametric Mann-Whitney-U-test), as some authors discuss stronger gender differences in the language of younger children (Beltz et al., 2013; Hayiou-Thomas et al., 2014; Lange et al., 2016; Ullman et al., 2008; Wallentin, 2009). There were no significant gender differences regarding the narrative comprehension total score, neither for monolingual (after retelling: $F(1, 198) = 0.352$, $p = .554$, after telling: $F(1, 198) = 0.643$, $p = .643$) nor for simultaneously bilingual children (after retelling: $F(1, 65) = 0.172$, $p = .680$, after telling: $F(1, 65) = 0.430$, $p = 0.514$). Monolingual boys and girls did not differ significantly in terms of the three subscores goal comprehension (after retelling: $F(1, 198) = 0.392$, $p = .531$, after telling: $F(1, 198) = 0.576$, $p = .449$), internal state comprehension (after retelling: $F(1, 198) = 3.076$, $p = .081$, after telling: $F(1, 198) = 0.086$, $p = .769$) and explanations (after retelling: $F(1, 198) = 0.433$, $p = .511$, after telling: $F(1, 198) = 0.001$, $p = .980$). Simultaneously bilingual boys and girls did not differ in terms of goal comprehension (after retelling: $F(1, 65) = 0.063$, $p = .803$, after telling: $F(1, 65) = 0.204$, $p = .653$), internal state comprehension (after retelling: $F(1, 65) = 1.197$, $p = .278$, after telling: $F(1, 65) = 2.041$, $p = .158$) and explanations (after retelling: $F(1, 65) = 0.000$, $p = .991$, after telling: $F(1, 65) = 0.004$, $p = .950$). Age group comparisons using Mann-Whitney-U-tests for monolingual and simultaneously bilingual children also showed no significant differences in the comprehension skills of boys and girls. However, as the age groups of the simultaneously bilingual boys and girls are very small, the results are only indicative.

5. Discussion and conclusion

This chapter described the development of narrative comprehension among monolingual and simultaneously bilingual German speaking children aged 4;6 to 5;11 who answered comprehension questions after telling and after retelling a story. In addition, the relationships between narrative comprehension and nonverbal cognitive skills as well as gender were investigated.

With regard to the development of narrative comprehension, significant differences were found between the three age groups (4;6–4;11, 5;0–5;5, 5;6–5;11) for monolinguals only. There was a strong effect of age group on narrative comprehension after both telling and retelling a story. This means that in the short period of about fourteen months, performance increases in narrative comprehension of monolinguals are observed (although children are 4;6 to 5;11 years old, average ages are: AG1 = 56.9 months, AG2 = 63.0 months, AG3 = 68.8 months). However, no significant increase in narrative comprehension is evident for the three age groups (post hoc-analyses), but for larger time intervals. The results of the monolingual children are in line with previous studies describing a development of narrative comprehension (Bohnacker, 2016; Hayward et al., 2009; Kapalková et al., 2016; Lindgren, 2018; Maviş et al., 2016; Otwinowska et al., 2018; Roch et al., 2016; Westerveld et al., 2012). For example, Lindgren (2018) noted similar increases in narrative comprehension score with age for monolingual and bilingual Swedish children. Considering that the maximal score for narrative comprehension was ten points, this ability seems to offer growth potential for older children (AG3: comprehension after retelling: monolinguals: $M = 6.9$, simultaneously bilinguals $M = 6.6$, AG3: comprehension after telling: monolinguals: $M = 5.6$, simultaneously bilinguals: $M = 5.0$).

In contrast to the results of monolinguals, the results for the simultaneously bilingual children in this study differ from those of several other studies (Bohnacker, 2016; Maviş et al., 2016; Roch et al., 2016). However, these studies focused on age groups that included children with more than six months age difference each (e.g. twelve months) and analyzed differences among groups with higher age separation (e.g. three-year-olds compared to six-year-olds), so that increasing performance becomes more visible. The age range of the children in the present study was broken down into six-month periods, which had the potential to show the development of narrative comprehension skills that would otherwise be overlooked by the more common one-year age range. In addition, previous studies monitored the languages of the participating children (e.g. only Swedish-English, Swedish-Turkish or Italian-English children were included). The bilinguals in these studies may therefore be less heterogeneous than the bilinguals in the present study. The present results show that the language abilities of the simultaneous bilinguals seem to be so different that there is no clear age trend in narrative comprehension. Despite a similar quantity of input (kindergarten attendance), similar evaluation of receptive and productive language skills by the parents, similar milestones in early language development (e.g. age of first word and first sentence in German), and similar results in terms of HASE and CPM (average standard deviation), no age trend for narrative comprehension can be determined for the simultaneous bilinguals.

Only a few monolingual and simultaneously bilingual children as young as age five could correctly answer all ten comprehension questions about goals, internal states and explanations as well as the meaning of the story as a whole. Although age-matched monolingual and bilingual groups strongly overlap in terms of narrative comprehension, the age effect is more pronounced for the monolinguals. Yet the monolinguals ($N = 199$) and simultaneous bilinguals ($N = 66$) were similar in terms of German proficiency evaluated by the parents and the results of the HASE: Both the monolinguals' and the bilinguals' parents rated their child's comprehension skills (both more than 90%) and production skills (both more than 91%) as "good". The results of HASEs subtests repetition of sentences, number sequences and non-words as well as the results of CPM regarding nonverbal cognitive skills did not show significant differences between the monolinguals and simultaneous bilinguals.

In the present study, there were significant age-independent differences between the narrative comprehension skills in the two elicitation modes for monolinguals and simultaneous bilinguals: The children scored more points answering comprehension questions after retelling (Cat) than after telling a story (Baby Birds). Listening to the story may have improved the child's narrative comprehension, as described in several studies (Isbell, Sobol, Lindauer, & Lowrance, 2004; Maviş et al., 2016; Otwinowska et al., 2018). When they had to answer the questions without listening to it beforehand, they had to draw conclusions from only studying the pictures. Another possibility might be that the comprehension questions of the two stories are of different difficulty, as has been argued by Bohnacker and Lindgren (in press) and Bohnacker, Lindgren, and Öztekin (2019) and Bohnacker, Öztekin, and Lindgren (this volume). These researchers, as well as Lindgren (2018), found that children achieved higher comprehension scores for Cat/Dog than for Baby Birds/Baby Goats, even though all stories were administered in the telling mode. The Cat story comprehension questions seem easier to answer than the questions on the Baby Birds story (Table 3), especially questions relating to episode three. As also shown by Bohnacker et al. (2019, this volume), in the third episode of Baby Birds, the child must see the dog as a human being (e.g. the dog wants to protect the birds from the cat and is proud of it afterwards), while in the third episode of Cat a typical animal goal must be set (the cat wants to eat the boy's fish) and a boy's reaction to that (feeling angry).

The order of the tasks may be another reason children performed better on the first narrative comprehension task, due to a higher attention level or greater interest in the new task format than on the second comprehension task where boredom or fatigue may have set in. It is also possible that the two story contents were of different interest to the children. For instance, the second episode of the Cat story (Table 3) allows the child to identify with the boy who loses his ball, a situation that preschoolers may have experienced themselves. What happens to a young boy (in Cat) may be easier to understand and be of greater interest to the children of this sample because it is closer to the children's world than to understand the goals, actions, and internal states of an animal, i.e. the dog in Baby Birds (see also Bohnacker et al., 2019, this volume). This explanation only applies to parts of the Cat story and thus to six comprehension questions that require taking on the boy's perspective. Further research on task effects seems to be necessary to evaluate these potential explanations.

The impact of nonverbal cognitive skills on narrative comprehension proved weak but significant for the monolinguals' overall narrative comprehension score after telling and retelling. One explanation for this finding might be that CPM measures *nonverbal* skills such as logical reasoning and MAIN narrative comprehension requires *verbal* skills to understand the questions and express answers, even if those answers need not to be grammatically or lexically perfect (Bohnacker et al., 2019, this volume; Bohnacker & Lindgren, in press). Both the monolinguals' and the simultaneous bilinguals' comprehension subscores for explanations (after telling a story) correlated significantly with nonverbal cognitive skills. Already Curenton (2010), Montgomery et al. (2009) and Murfett et al. (2008) pointed out links between cognitive skills and narrative comprehension, which were partially confirmed in this sample for nonverbal skills. Because CPM measures fluid intelligence, including skills such as inductive and logical thinking, and sequential reasoning (Mickley & Renner, 2010), it is more likely to reveal a correlation with comprehension scores after telling than comprehension after retelling, especially with explanation questions and less likely with goal questions and internal state questions or the total score. To answer explanation questions about the reasons for internal states of the characters without having listened to the story beforehand may require more focused reasoning and logical thinking, as the child alone has to draw inferences and conclusions by looking only at the pictures. The results show that there is a link between nonverbal cognition skills and narrative comprehension, although there may be stronger impact factors (low correlation coefficients). This means that the ability to answer (explanation) questions about a story builds in part on inductive and logical thinking and sequential reasoning.

For the simultaneous bilinguals, only one significant correlation was found, namely between nonverbal cognitive skills and explanation questions after telling a story. One explanation for this could be the heterogeneity of the group due to different language combinations, differences in linguistic distance between the respective languages and German, and variation in quality and quantity of German input. Considering these aspects, a group of 66 simultaneous bilinguals may have been quite small. Another explanation for the lack of associations could be that the narrative comprehension performance of the simultaneous bilinguals is more related to specific receptive language skills to understand the questions, and productive language skills to answer them adequately, than to nonverbal cognitive skills. Further research on factors that impact on narrative comprehension, such as children's German receptive and expressive vocabulary skills, is needed to examine this possibility.

There were no significant gender differences for the children's narrative comprehension. This result is in line with the results of John et al. (2003) and Maviş et al. (2016), but is in contrast to the results of Aldrich & Brooks (2017). One explanation for different results could be the assessment procedure, including the child's task. While comprehension studies that find gender differences tend to assess personal stories that are based on gender differences in experiences (Nicolopoulou, 2008), studies such as those using MAIN elicit fictional stories with tightly controlled stimulus material and procedures (Maviş et al., 2016; Kunnari et al., 2016), and thus the children may have less leeway to tell divergent stories (Bohnacker & Lindgren, in press; Bohnacker et al., 2019, this volume). Other reasons for divergent results could be differences in sample size, age and languages, as well as different instruments for assessing narrative skills (Matthews et al., 2018). The potential impact of gender on other narrative abilities thus remains to be determined, such as the use of internal state terms when telling a story.

In summary, from age 4;6 to age 5;11, a stronger development of narrative comprehension skills could be demonstrated for monolinguals than for bilinguals. No effects of gender and only weak effects of nonverbal cognitive skills were found. Caution is advised before generalizing these findings, since the present study has several methodological limitations: The group of simultaneous bilinguals was quite heterogeneous, with many different language combinations; moreover, missing data reduced the number of participants. Also, stories and elicitation modes were not counterbalanced, as Cat (retelling) was always administered first, followed by Baby Birds (telling). If the data collection had taken place with both stories in alternating order, it would have been possible to draw more accurate conclusions about the development of narrative comprehension skills.

In the future the role of internal and external factors should be explored to better explain variation in the narrative comprehension skills of monolingual and bilingual children.

Acknowledgements

I thank the volume editors, Ute Bohnacker and Natalia Gagarina, for generously providing me with ideas how to improve the discussion, as well as for their helpful feedback on earlier drafts of the entire chapter.

References

Aldrich, N. J., & Brooks, P. J. (2017). Linguistic and socio-cognitive predictors of school-age children's narrative evaluations about jealousy. *First Language*, 37(2), 130–149. https://doi.org/10.1177/0142723716679797

Anderson, R. (1994). Role of reader's schema in comprehension, learning and memory. In R. Ruddell, M. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 469–482). Newark, DE: International Reading Association.

Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children. Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137

Beltz, A. M., Blakemore, J. E. O. & Berenbaum, S. A. (2013). Sex differences in brain and behavioral development. In H. Tager-Flusberg, P. Rakic, & J. Rubenstein (Eds.), *Comprehensive developmental neuroscience, Vol. 3: Neural circuit development and function in the healthy and diseased brain* (pp. 467–499). Oxford: Elsevier. https://doi.org/10.1016/B978-0-12-397267-5.00064-9

Berman, R. A., & Slobin, D. I. (Eds.). (1994). *Relating events in narrative: A crosslinguistic developmental study*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bildungsberichterstattung (2018). *Bildung in Deutschland 2018. Ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung*. Bielefeld: wbv media.

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

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Bohnacker, U., & Lindgren, J. (in press). MAIN story comprehension: What can we expect of a typically developing child? In S. Armon-Lotem & K. Grohmann (Eds.), *LITMUS in action: Cross-comparison studies across Europe*. Amsterdam: John Benjamins. Preprint retrieved from <http://uu.diva-portal.org/smash/get/diva2:1348899/FULLTEXT01.pdf> (23 June, 2020).

Bohnacker, U., Lindgren, J., & Öztekin, B. (2019). Understanding MAIN picture stories: A study of Swedish-Turkish children aged 4–7. Oral presentation at Bilingual Acquisition of Language and Literacy (BiALL) conference, 24 May 2019, Leibniz-ZAS, Berlin.

Bohnacker, U., Öztekin, B., & Lindgren, J. (this volume). Bilingual Turkish-Swedish children's understanding of MAIN picture sequences: Individual variation, age, language and task effects. In U. Bohnacker & N. Gagarina (Eds.), *Developing narrative comprehension: Multilingual Assessment Instrument for Narratives*. Amsterdam: John Benjamins. https://doi.org/10.1075/sibil.61.04boh

Bohn-Gettler, C. M., Rapp, D. N., van den Broek, P., Kendeou, P. & White, M. J. (2011). Adults' and children's monitoring of story events in the service of comprehension. *Memory & Cognition*, 39(6), 992–1011. https://doi.org/10.3758/s13421-011-0085-0

Brown, D. D., Lile, J., & Burns, B. M. (2011). Basic language skills and young children's understanding of causal connections during storytelling. *Reading Psychology*, 32(4), 372–394. https://doi.org/10.1080/02702711.2010.495573

Burris, S. E. & Brown, D. D. (2014). When all children comprehend: increasing the external validity of narrative comprehension development research. *Frontiers in Psychology*, 5(168), 1–16.

Curenton, M. (2010). Understanding the landscapes of stories: The association between preschoolers' narrative comprehension and production skills and cognitive abilities. *Early Child Development and Care*, 181(6), 791–808. https://doi.org/10.1080/03004430.2010.490946

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin: Springer. https://doi.org/10.1007/978-3-642-41089-5

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language, and Hearing Research*, 47(6), 1301–1318. https://doi.org/10.1044/1092-4388(2004/098)

Florit, E., Roch, M., & Levorato, M. C. (2011). Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. *Discourse Processes*, 48(2), 119–138. https://doi.org/10.1080/0163853X.2010.494244

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics, 56*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–276). Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137-011

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics, 63*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Gardner-Neblett, N., & Sideris, J. (2018). Different tales: The role of gender in the oral narrative – Reading link among African American children. *Child Development*, 89(4), 1328–1342. https://doi.org/10.1111/cdev.12803

Gerson, S. A., & Woodward, A. L. (2013). The goal trumps the means: Highlighting goals is more beneficial than highlighting means in means-end training. *Infancy*, 18(2), 289–302. https://doi.org/10.1111/j.1532-7078.2012.00112.x

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, 27(3), 689–705. https://doi.org/10.1017/S0305000900004414

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2014). Language impairment from 4 to 12 years: Prediction and etiology. *Journal of Speech, Language, and Hearing Research*, 57(3), 850–864. https://doi.org/10.1044/2013_JSLHR-L-12-0240

Hayward, D. V., Schneider, P., & Gillam, R. B. (2009). Age and task-related effects on young children's understanding of a complex picture story. *The Alberta Journal of Educational Research*, 55(1), 54–72.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104(1), 53–69. https://doi.org/10.1037/0033-2909.104.1.53

Isbell, R., Sobol, J., Lindauer, L., & Lowrance, A. (2004). The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early Childhood Education Journal*, 32(3), 157–163. https://doi.org/10.1023/B:ECEJ.0000048967.94189.a3

John, S. F., Lui, M., & Tannock, R. (2003). Children's story retelling and comprehension using a new narrative resource. *Canadian Journal of School Psychology*, 18(1–2), 91–113. https://doi.org/10.1177/082957350301800105

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The Index of Narrative Microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology*, 15(2), 177–191. https://doi.org/10.1044/1058-0360(2006/017)

Kapalková, S., Polisenská, K., Marková, L., & Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak-English children: A cross-language comparison. *Applied Psycholinguistics*, 37(1), 145–164. https://doi.org/10.1017/S0142716415000454

Karasinski, C., & Ellis Weismer, S. (2010). Comprehension of inferences in discourse processing by adolescents with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 53(5), 1268–1279. https://doi.org/10.1044/1092-4388(2009/09-0006)

Kendeou, P., Bohn-Gettler, C., White, M., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, 31(3), 259–272. https://doi.org/10.1111/j.1467-9817.2008.00370.x

Kim, Y.-S. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, 86(1), 128–144. https://doi.org/10.1111/cdev.12293

Kim, Y.-S. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology*, 141, 101–120. https://doi.org/10.1016/j.jecp.2015.08.003

Kunnari, S., Välimaa, T., & Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Applied Psycholinguistics*, 37(1), 123–144. https://doi.org/10.1017/S0142716415000442

Lange, B. P., Euler, H. A., & Zaretsky, E. (2016). Sex differences in language competence of 3- to 6-year-old children. *Applied Psycholinguistics*, 37(1), 1417–1438. https://doi.org/10.1017/S0142716415000624

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research*, 38(2), 415–425. https://doi.org/10.1044/jshr.3802.415

Lindgren, J. (2018). *Developing narrative competence. Swedish, Swedish-German and Swedish-Turkish children aged 4–6* (Studia Linguistica Upsaliensia 19). Uppsala: Acta Universitatis Upsaliensis.

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology*, 29(4), 327–365. https://doi.org/10.1080/02702710802165416

Lynch, J. S., & van den Broek, P. (2007). Understanding the glue of narrative structure: Children's on-and off-line inferences about characters' goals. *Cognitive Development*, 22(3), 323–340. https://doi.org/10.1016/j.cogdev.2007.02.002

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maital, S. L., Dromi, E., Sagi, A. & Bornstein, M. H. (2000). The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language*, 27(1), 43–67. https://doi.org/10.1017/S0305000999004006

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition, and executive functions. *Language Learning and Development*, 14(3), 186–223. https://doi.org/10.1080/15475441.2018.1455584

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, 37(1), 69–89. https://doi.org/10.1017/S0142716415000429

Meisels, S. J., Marsden, D. B., Wiske Stone, M., & Henderson, L. W. (1997). *Examiner's manual: Early screening inventory-revised*. Ann Arbor, MI: Rebus.

Melzer, J., Rißling, J.-K., & Petermann, F. (2016). Kognitive Kompetenzen und Sprachentwicklung bei Kindern im Alter zwischen vier und fünf Jahren. *Zeitschrift für Neuropsychologie*, 27(1), 37–51. https://doi.org/10.1024/1016-264X/a000170

Mickley, M., & Renner, G. (2010). Intelligenztheorie für die Praxis: Auswahl, Anwendung und Interpretation deutschsprachiger Testverfahren für Kinder und Jugendliche auf Grundlage der CHC-Theorie. *Klinische Diagnostik und Evaluation*, 3(4), 447–466.

Montgomery, J., Poluenko, A., & Marinellie, S. A. (2009). Role of working memory in children's understanding spoken narrative: A preliminary investigation. *Applied Psycholinguistics*, 30(3), 485–509. https://doi.org/10.1017/S0142716409090249

Murfett, R., Powell, M. B., & Snow, P. C. (2008). The effect of intellectual disability on the adherence of child witnesses to a "story grammar" framework. *Journal of Intellectual & Developmental Disability*, 33(1), 2–11. https://doi.org/10.1080/13668250701829811

Nelson, K. (2010). Developmental narratives of the experiencing child. *Child Development Perspectives*, 4(1), 42–47. https://doi.org/10.1111/j.1750-8606.2009.00116.x

Nicolopoulou, A. (2008). The elementary forms of narrative coherence in young children's storytelling. *Narrative Inquiry*, 18(2), 299–325.

Nicolopoulou, A., Brockmeyer, C., de Sá, A., & Ilgaz, H. (2014). Narrative performance, peer group culture, and narrative development in a preschool classroom. In A. Cekaite, S. Blum-Kulka, V. Grover, & E. Teubal (Eds.), *Children's peer talk: Learning from each other* (pp. 42–62). New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139084536.006

Orrantia, J., Múñez, D., & Tarín, J. (2014). Connecting goals and actions during reading: the role of illustrations. *Reading and Writing*, 27(1), 153–170. https://doi.org/10.1007/s11145-013-9437-4

Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 1–25. https://doi.org/10.1080/13670050.2018.1434124
Pavlenko, A. (2008). Narrative analysis. In L. Wei & M. G. Moyer (Eds.), *The Blackwell guide to research methods in bilingualism and multilingualism* (pp. 311–325). Malden, MA: Blackwell.
Peterson, C., & McCabe, A. (1991). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York, NY: Plenum Press.
Quasthoff, U., & Stude, J. (2018). Narrative Interaktion: Entwicklungsaufgabe und Ressource des Erzählerwerbs. *Zeitschrift für Literaturwissenschaft und Linguistik*, 48(2), 249–275. https://doi.org/10.1007/s41244-018-0092-8
Raven, J. C., Bulheller, J., & Häcker, J. H. (2002). *Coloured Progressive Matrices (CPM)*. Frankfurt: Pearson.
Renner, G., & Mickley, M. (2015). Intelligenzdiagnostik im Vorschulalter: CHC-theoretisch fundierte Untersuchungsplanung und Cross-battery-assessment. *Frühförderung interdisziplinär*, 34(2), 67–82. https://doi.org/10.2378/fi2015.art07d
Ringmann, S. (2014). Therapie der Makrostruktur von Erzählungen. *Sprachförderung und Sprachtherapie in Schule und Praxis*, 3(3), 147–155.
Ringmann, S. (2013). Therapie der Erzählfähigkeit. In S. Ringmann & J. Siegmüller (Eds.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen. Schuleingangsphase* (pp. 163–187). Munich: Elsevier.
Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian-English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, 37(1), 49–67. https://doi.org/10.1017/S0142716415000417
Schmitter-Edgecombe, M., & Creamer, S. (2010). Assessment of strategic processing during narrative comprehension in individuals with mild cognitive impairment. *Journal of the International Neuropsychological Society*, 16(4), 661–671. https://doi.org/10.1017/S1355617710000433
Schöler, H., & Brunner, M. (2008). *Heidelberger auditives Screening in der Einschulungsuntersuchung (HASE)*. Binswangen: Westra.
Sommerville, J. A., & Woodward, A. L. (2005). Infants' sensitivity to the causal features of means–end support sequences in action and perception. *Infancy*, 8(2), 119–145. https://doi.org/10.1207/s15327078in0802_2
Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives* (pp. 53–120). Norwood, NJ: Ablex.
Szagun, G. (2006). *Sprachentwicklung beim Kind*. Weinheim: Beltz.
Szaflarski, J. P., Altaye, M., Rajagopal, A., Eaton, K., Meng, X., Plante, E. & Holland, S. K. (2012). A 10-year longitudinal fMRI study of narrative comprehension in children and adolescents. *Neuroimage*, 63(3), 1188–1195. https://doi.org/10.1016/j.neuroimage.2012.08.049
Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, 15(3), 249–275. https://doi.org/10.1080/01638539209544812
Trabasso, T., van den Broek, P., & Liu, L. (1988). A model for generating questions that assess and promote comprehension. *Questioning Exchange*, 2(1), 25–38.
Tsimpli, I. M., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics*, 37(1), 195–216. https://doi.org/10.1017/S0142716415000478

Ullman, M. T., Miranda, R. A., & Travers, M. (2008). Sex differences in the neurocognition of language. In J. B. Becker, K. J. Berkley, N. Geary, E. Hampson, J. Herman, & E. Young (Eds.), *Sex on the brain: From genes to behavior* (pp. 291–309). New York, NY: Oxford University Press.

Van den Broek, P., Lynch, J. S., Naslund, J., Ievers-Landis, C. E., & Verduin, K. (2003). The development of comprehension of main ideas in narratives: Evidence from the selection of titles. *Journal of Educational Psychology*, 95(4), 707–718. https://doi.org/10.1037/0022-0663.95.4.707

Van Kleeck, A., Lange, A., & Schwarz, A. L. (2011). The effect of race and maternal education level on children's retells of the Renfrew Bus Story-North American Edition. *Journal of Speech, Language, and Hearing Research*, 54(6), 1546–1561. https://doi.org/10.1044/1092-4388(2011/10-0079)

Vygotsky, L. (1962). Thought and word. In L. Vygotsky, E. Hanfmann, & G. Vakar (Eds.), *Thought and language* (pp. 119–153). Cambridge, MA: The MIT Press. https://doi.org/10.1037/11193-007

Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain & Language*, 108(3), 175–183. https://doi.org/10.1016/j.bandl.2008.07.001

Westerveld, M. F., Gillon, G. & Boyd, L. (2012). Evaluating the clinical utility of the Profile of Oral Narrative Ability for 4-year-old children. *International Journal of Speech-Language Pathology*, 14(2), 130–140. https://doi.org/10.3109/17549507.2011.632025

Westby, C. E. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and reading disabilities* (pp. 157–232). Boston, MA: Allyn & Bacon.

Bilingualism effects in the narrative comprehension of children with Developmental Language Disorder and L2-Greek

Links with language, executive function and Theory of Mind

Eleni Peristeri, Maria Andreou, Ianthi Maria Tsimpli and Stephanie Durrleman
University of Thessaly, Greece / University of Cologne, Germany / University of Cambridge, United Kingdom / University of Geneva, Switzerland

Narrative comprehension is a complex process that requires the ability to integrate language information from the speech signal with visual and contextual knowledge, while drawing also from social cognition and executive functions. Although many studies have examined narrative comprehension in typically-developing (TD) bilingual children and suggested a bilingual advantage, evidence for bilingualism effects in the narrative comprehension performance of children with Developmental Language Disorder (DLD) is scant. This study explores narrative comprehension of thirty 6 to 8 year old monolingual Greek and Albanian-Greek bilingual children with DLD, along with two groups of age-matched TD monolingual Greek and Albanian-Greek bilingual children. Children's narrative comprehension was assessed through the Greek versions of two stories (*Cat* and *Dog*) which have been designed for retelling within the Multilingual Assessment Instrument for Narratives tool (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012) of the COST Action IS0804. The children's language, Theory of Mind (ToM) and updating skills were independently measured. Both groups with DLD had lower language and executive function performance than TD children. Bilinguals with and without DLD, however, scored higher in narrative comprehension than their TD and DLD monolingual peers. Similarly, bilingual children with DLD outperformed their monolingual peers with DLD on the ToM task. TD children's narrative comprehension was predicted by their language and executive function performance, while DLD bilingual children's narrative comprehension was predicted by performance on the ToM

https://doi.org/10.1075/sibil.61.10per

task and their dominance in L2/Greek. The overall results indicate advantages for bilingual children with DLD in narrative comprehension and ToM, while suggesting a link between these enhanced skills.

Keywords: Developmental Language Disorder, bilingualism, narrative comprehension, executive functions, Theory of Mind

1. Introduction

Over the past decade, the positive effects of bilingualism on the language and cognitive development of children with Developmental Language Disorder (DLD) have come into sharper focus. Special emphasis has been placed on DLD children's narrative performance and the way bilingualism influences microstructural and macrostructural aspects of their narratives. Bilingual children with DLD have been reported to outperform their monolingual peers with DLD on narrative macrostructure, and more specifically on story structure complexity in narrative production (Tsimpli, Peristeri, & Andreou, 2016). Research in the way bilingualism affects microstructural properties of narrative production in children with DLD converges on the finding that typically developing (TD) bilingual children outperform their bilingual peers with DLD on a range of microstructural features, including vocabulary and morphosyntax (Altman, Armon-Lotem, Fichman, & Walters, 2016; Rezzonico, Chen, Cleave, Greenberg, Hipfner-Boucher, Johnson, & Girolametto, 2015; Tsimpli et al., 2016). On the other hand, cognitive skills in bilingual children with DLD have received limited attention as shown by the very small number of research studies. Some of these studies provide converging evidence that bilingualism contributes to better executive functions and mentalizing abilities in DLD (Peristeri, Baldimtsi, Tsimpli, & Durrleman, 2019; Tsimpli, Peristeri, & Andreou, 2017).

While positive effects of bilingualism in DLD children's narrative *production* and cognitive skills have been reported, it remains to be determined whether narrative *comprehension* in bilingual children with DLD is boosted and whether this boost would be mediated by linguistic proficiency, or by cognitive processes which have been reported to be positively affected by bilingualism, such as Theory of Mind (ToM) and executive functions (Peristeri et al., 2019; Tsimpli, Peristeri, & Andreou, 2017). The present study thus aims to investigate whether bilingualism indeed improves narrative comprehension, and whether this contribution is mediated by enhanced executive functions, better language and/or ToM skills in children with DLD. To this end, we evaluated the narrative comprehension performance of four groups of children; two groups of 6 to 8 year old monolingual Greek and bilingual Albanian-Greek-speaking children with a diagnosis of DLD, and two groups of age-matched TD monolingual Greek and bilingual Albanian-Greek children. Children's narrative comprehension was assessed in their L2/Greek with the Multilingual Assessment Instrument for Narratives (MAIN) (Gagarina, Klop, Kunnari, Tantele, Välimaa, Balčiūnienė, Bohnacker, & Walters, 2012), which was developed within the COST Action IS0804 "Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment". This COST Action addressed the issue of disentangling bi/multilingualism from language impairment and of profiling the language and cognitive skills of bilingual children with DLD.

Narrative comprehension constitutes an important part of the assessment procedure of MAIN. The set of comprehension questions focuses on macrostructure components and internal state terms, and it is administered after the narrative production part of the assessment procedure. MAIN includes ten comprehension questions, three targeting goals in episodes, six targeting internal states as goals, initiating events or reactions, and the final question, as well as some of the earlier questions, requiring children to infer information that is not explicitly mentioned in the narrative input provided for the retelling mode.

MAIN has been successfully used in a wide body of research on TD bilingual children of different language combinations to assess narrative comprehension. This research has revealed better comprehension in response accuracy on comprehension questions after having to tell a story based on a model as compared to when they had to tell a story without a model (Maviş, Tunçer, & Gagarina, 2016). These results, from Turkish-German bilinguals, seem in line with the findings of Otwinowska, Mieszkowska, Białecka-Pikul, Opacki, and Haman's study (2018) with Polish-English bilinguals, whose narratives were boosted by retelling regardless of the language of narration (Polish, English). In Bohnacker's (2016) story-telling and comprehension study, 5 year old TD bilingual Swedish-English children were found to score lower than 6 to 7 year old Swedish-English bilinguals in narrative comprehension, and both age groups were found to struggle when answering questions requiring inferring internal states (thoughts, feelings) as reactions beyond the purely physical and explicitly depicted (Bohnacker, 2016, p. 43). MAIN has also been successfully used for exploring narrative comprehension in one study involving children with DLD, that of Boerma, Leseman, Timmermeister, Wijnen, and Blom (2016), who examined the narrative comprehension skills of 5 to 6 year old monolingual and bilingual children matched on exposure to L2/Dutch before age 4 and current exposure to Dutch at home. This specific study focused on children's performance across MAIN's three types of comprehension questions. The findings did not reveal any bilingualism effect in DLD children's comprehension performance across the three taxonomies relative to their monolingual peers with DLD. The overall findings from this body of work suggest that bilingualism does not create a disadvantage in the narrative comprehension of either TD or DLD children. However, the evidence for bilingualism effects in children with DLD is sparse, and the specific links between narrative comprehension and children's language ability, executive functions and ToM skills is inexistent. More work is thus needed to clarify the factors underlying narrative comprehension in bilingual children with DLD.

1.1 Narrative comprehension and language ability

Recently, narrative comprehension in children has been tested through the lens of language demands, and more specifically, children's vocabulary skills and grammatical knowledge. Comprehension of oral narratives imposes clear demands on language; in order for the listener to be able to construct a coherent story, s/he has to comprehend words, structures and subsequently derive propositional meanings. Language demands on listening comprehension have been usually measured through children's vocabulary, syntactic comprehension and sentence repetition abilities. Studies have revealed consistent patterns of medium-to-large correlations between listening comprehension and language skills in TD children (Daneman & Blennerhassett, 1984; Florit, Roch, & Levorato, 2011, 2013; Kidd, 2013; Kim, 2016; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012), as well as in children with DLD (Karasinski & Ellis Weismer, 2010; Tsimpli et al., 2016). These findings imply that establishing a true effect of bilingualism on DLD children's narrative comprehension skills requires measuring their language ability.

1.2 Narrative comprehension and executive functions

Apart from language skills, other higher-level cognitive processes have been shown to play a role in children's comprehension, such as working memory, monitoring or maintenance of information in memory, and inference-making skills (Chrysochoou & Bablekou, 2011; Florit, Roch, Altoe, & Levorato, 2009; Kim, 2016; Strasser & Francisca del Río, 2014; Was & Woltz, 2007). These processes can be understood to belong to executive functions. While definitions of executive functioning may vary due to its multi-faceted nature, there is general consensus that it is dependent on a number of subskills, such as inhibition, attention and working memory. Especially, verbal working memory has been found to be more likely than other executive functions to affect children's listening comprehension especially when they are required to infer implicit knowledge that guides their expectations while listening to a story (Kim, 2016), or during comprehension monitoring and mismatch detection (Oakhill, Hartt, & Samols, 2005).

Children with DLD have been systematically shown to exhibit deficits in a wide range of executive functions, including working memory, inhibition, and attention (Lukács, Ladányi, Fazekas, & Kemény, 2016; Marton, 2008; Marton, Campanelli, Eichorn, Scheuer, & Yoon, 2014; see Vissers, Koolen, Hermans, Scheper, & Knoors, 2015 for a review). Since narrative comprehension is thought to depend on specific aspects of executive functioning, limitations to DLD children's higher-level cognitive abilities have been shown to shape their narrative performance in a number of studies. For instance, children with DLD in Dodwell and Bavin's (2008) narrative study were able to generate information that was difficult for them to infer in narrative comprehension. The asymmetry between children's narrative production and comprehension regarding contextually-based inferences has been attributed to DLD children's limited memory storage and retrieval capacity. Difficulties in integrating context-dependent cues due to limited memory and information processing capacities has also been reported by Ford and Milosky (2003), who found that children with DLD were less proficient at making accurate inferences regarding the characters' emotions in orally-presented picture stories, in spite of the fact that the same children were able to verbally label these emotional expressions in a picture-naming task. Blom and Boerma (2016) also measured verbal short-term and working memory, and sustained attention in TD children and children with DLD, and found significant correlations between story comprehension and all cognitive measures for children with DLD, while for TD children comprehension was found to correlate with verbal short-term memory only (Blom & Boerma, 2016: 307).

Few studies have considered executive functions in bilingual children with DLD and the evidence is rather mixed. Engel de Abreu, Cruz-Santos, and Puglisi (2014) showed that bilingual children with DLD did not differ from either TD monolingual or bilingual children on visuospatial working memory measures; yet, the bilingual group with DLD performed less well than TD groups on tasks tapping into selective attention and interference suppression. Finally, Laloi, de Jong, and Baker (2017) found that both bilingual and monolingual children with DLD performed similarly to their TD peers on the accuracy measure of a response inhibition task, yet both groups with DLD exhibited significantly longer reaction times in their responses.

1.3 Narrative comprehension and Theory of Mind

ToM skills have been associated with narrative comprehension in previous research (Atkinson, Slade, Powell, & Levy, 2017; Kim, 2016). ToM refers to the ability to mentalize from one's perspective in relation to others' mental states, i.e. intentions, desires, emotions and beliefs, and to predict, describe and explain resulting behaviors. Interpreting descriptions about other individuals' actions relies on an understanding of the individuals' inner state. In the narrative genre in particular, coherence relations among events mainly depend on effectively inferring the characters' goals and motivations (Lynch & van den Broek, 2007, p. 337). Thus, having difficulty recognizing others as beings who are guided by mental states would lead to weaker anticipation of perspective-appropriate interpretations of discourse. Studies on the contribution of ToM to children's narrative comprehension report mixed results. One study (Strasser & del Rio, 2014) with TD kindergarten children found no role for mentalizing in explaining variance on a wordless-book comprehension task, once children's language ability, executive functions, inferencing and comprehension monitoring skills were controlled for. On the other hand, Kim's (2016) listening comprehension study with a large sample of first grade TD children found large correlations between ToM and listening comprehension of short stories, even when children's grammatical knowledge, working memory and inferencing skills were taken into account. The overall findings suggest that TD children's listening comprehension may be captured by ToM skills, yet, the extent to which ToM mediates narrative comprehension in children with DLD remains underexplored.

Bilingual children with DLD provide an intriguing test case regarding the (potentially compensatory) role of ToM skills in narrative comprehension, all the more since bilingualism has been shown to positively affect ToM skills in both TD children and children with DLD. More specifically, TD bilingual children have been shown to outperform their monolingual peers on tests assessing false belief understanding (Bialystok & Senman, 2004; Goetz, 2003). According to Dennet (1978), tests of false belief are the most reliable method of assessing ToM because they imply that one's mental representation of reality and a given reality may dissociate. Tsimpli and colleagues (2017) have recently found that bilingual children with DLD outperformed their monolingual peers with DLD on verbal second-order false belief attribution tasks, and more importantly, that their ToM abilities were a significant predictor of referentially appropriate pronoun use in a short discourse production task.

The overall results suggest that bilingualism positively affects DLD children's sensitivity to others' perspectives in discourse production. While it is intuitively plausible that ToM skills are called upon for narrative comprehension, the ToM tasks used so far have either tapped onto advanced vocabulary or required children to comprehend stories of around 200 words each (e.g. Gillott, Furniss, & Walter, 2004; Ziatas, Durkin, & Pratt, 1998); it is thus possible that monolingual children with DLD had difficulty performing the ToM tasks due to a deficit in accessing the lexical or/and grammatical content of the stimuli in the tasks, and not due to a ToM deficit per se (Miller, 2001). Non-verbal ToM tasks may therefore provide more refined measures of mentalizing skills for children with DLD whose phenotype has been extensively shown to be marked by deficits in language.

2. Aims and research questions

The present study seeks to investigate narrative comprehension of thirty 6 to 8 year old monolingual Greek and Albanian-Greek bilingual children with DLD, along with two groups of age-matched TD monolingual Greek and Albanian-Greek bilingual children. Children's narrative comprehension was assessed through the Greek versions of two stories (*Cat* and *Dog*) which have been designed for retelling within the COST Action IS0804 (Gagarina et al., 2012; see the Appendix for the story texts in Greek and their corresponding translations in English). All testing was carried out in Greek in the absence of standardized language screening and assessment measures in bilingual children's L1/Albanian. Narrative data were also collected in Greek, the children's L2, due to the authors' lack of knowledge of Albanian that could have otherwise allowed the administration of the narrative task, the transcription and the analysis of the narrative data in Albanian.

All participants were administered two language ability tasks that tested expressive vocabulary and sentence repetition skills. These tasks allowed us to examine whether potential differences between groups in narrative comprehension could be due to the children's language ability, and whether language ability takes precedence over higher cognitive skills, such as executive functions or/and ToM, in the narrative comprehension performance of monolingual and bilingual TD children and children with DLD.

Participants were also tested on an executive function task, specifically, a non-verbal 2-back task in which they were required to remember if the digit they saw on the screen was the same as the one presented two positions back in a digit sequence. The 2-back task has been claimed to implicate a broad range of executive functions, including working memory, inhibition and updating (Friedman & Miyake, 2004). To perform successfully, the child has to maintain digit sequences in her/his working memory, suppress non-relevant, interfering digit information, and retain the activation levels of target digits.

Finally, monolingual and bilingual children with DLD were administered a ToM task that measured their ability to reason about an external agent's false belief. The ToM task was non-verbal so as to avoid the possibility that the children's performance would be affected by their lower language proficiency stemming from DLD and/or bilingualism. To the best of our knowledge, this is the first study to compare narrative comprehension in monolingual and bilingual children with DLD in Greek using MAIN (Gagarina et al., 2012), and also the first to document (possible) interactions between narrative comprehension and language, executive functions, and ToM skills across the two groups.

We thus formulated the research questions and hypotheses of the study as follows:

Question 1. What is the effect of DLD, the effect of bilingualism and their combined effect on narrative comprehension?

Hypothesis 1. Based on Boerma and colleagues' (2016) findings, we first hypothesized that a negative effect of DLD would be found on the narrative comprehension measure, while bilingualism was not expected to create a disadvantage in the narrative comprehension of either TD children or children with DLD. As such, we expected bilingual children with and without DLD to perform similarly to their monolingual peers.

Question 2. Does updating measured in a non-verbal executive function task affect narrative comprehension in monolingual and bilingual TD children and children with DLD?

Hypothesis 2. Based on previous research (e.g. Bialystok & Craik, 2010), we assumed a positive effect of bilingualism on both TD and DLD children's updating skills, which would also facilitate children's narrative comprehension.

Question 3. Does the ToM performance of bilingual children with DLD differ from monolingual children with DLD?

Hypothesis 3. Based on previous findings according to which bilingualism positively affects ToM skills in children with DLD (Tsimpli et al., 2017), we hypothesized that the bilingual group with DLD would outperform the monolingual group in the ToM measure.

Question 4. Which of the examined factors, namely, language ability, updating skills, and ToM, best captures narrative comprehension in monolingual and bilingual TD and DLD children? Is the contribution of the aforementioned factors to narrative comprehension comparable across the four groups?

Hypothesis 4. Based on previous evidence (Kim, 2016; Oakhill et al., 2005) that language ability, executive function and ToM skills mediate narrative comprehension in TD children, we hypothesized that both language and higher-order cognitive skills would be closely linked with TD monolingual and bilingual children's comprehension in the narrative task. We further hypothesized that the influence of language on narrative comprehension would be weaker for both monolingual and bilingual children with DLD as compared to their TD peers due to the former groups' language impairment. Finally, we predicted that, if bilingualism improves updating and ToM skills in children with DLD as compared to their monolingual peers, then bilingual children with DLD would benefit from their updating and mentalizing capacity, which would in turn effectively support the challenges associated with narrative comprehension.

3. Method

3.1 Participants

A total of one hundred and twenty 6 to 8 year old children divided into 4 groups participated in this study: 30 monolingual Greek-speaking children with DLD (DLD-Mono); 30 Greek-Albanian children with DLD (DLD-Bi); 30 TD monolingual Greek-speaking children (TD-Mono); and 30 Greek-Albanian TD children (TD-Bi). The children were matched across groups for gender. There were no significant differences between groups in age, F (3, 116) = .381, p = .767, η^2 = .06. Furthermore, the children's non-verbal intelligence (or else, performance IQ/PIQ) was above clinical levels of intellectual impairment, as measured through the percentile scores on the Greek version of the Wechsler Intelligence Scale for Children-Revised (WISC-III) (Wechsler, 1992; adapted in Greek by Georgas, Paraskevopoulos, Besevegis, Giannitsas, & Mylonas, 2003). There was no significant difference in PIQ percentiles among the four groups, F (3, 116) = .205, p = .341, η^2 = .09 (see Table 1).

Table 1. Groups' mean age and PIQ percentile scores (and *SD*s)

Groups	**Mean age (*SD*) Age range**	**Mean PIQ percentile scores (*SD*)**
TD-Mono (N = 30)	6;9 (*0;6*) 6;1–7;9	50.5 (*11.8*)
TD-Bi (N = 30)	7;0 (*0;5*) 6;0–7;9	51.2 (*12.4*)
DLD-Mono (N = 30)	6;9 (*0;6*) 6;0–8;1	49.9 (*10.4*)
DLD-Bi (N = 30)	7;0 (*0;6*) 6;1–7;9	48.2 (*11.4*)

Note: PIQ: Performance IQ; TD-Mono: monolingual typically developing children; TD-Bi: bilingual typically developing children; DLD-Mono: monolingual children with Developmental Language Disorder; DLD-Bi: bilingual children with Developmental Language Disorder; *SD*: standard deviation

Typically-developing children were recruited from mainstream schools in Greece and they were included in the study if they had normal hearing and no speech, emotional or behavior problems, and no neurological or severe articulation/phonological deficits. The typically-developing children's profile was confirmed by information from health screening protocols, which were implemented prior to data collection as part of the Governmental Public Health Policy in Greek public education, and teachers' and parents' reports. Experimental data were collected following all children's parents' written consent, children's assent and obtainment of approval from the Research Ethics Committee of the Greek Ministry of Education.

Children with DLD, both monolingual and bilingual, were recruited from public diagnostic centers in Greece. In line with DSM-5 criteria (American Psychiatric Association, 2013), the children already had a speech and language therapist's/clinician's diagnosis of DLD in the absence of any hearing loss, autism, obvious neurological dysfunctions or motor deficits. The diagnosis of DLD was further supported by questionnaires, as well as language and neuropsychological testing. More specifically, parental questionnaires and language unit class teachers' reports confirmed significant delays in the children's early language milestones as well as expressive difficulties in both the oral and the written modality (Leonard, 1998). Both monolingual and bilingual children with DLD were administered the Diagnostic verbal IQ test for Greek school-aged children (Stavrakaki & Tsimpli, 2000), which includes a series of tasks that test morpho-syntactic production and comprehension skills. Bilingual children with DLD were found to score at least 2.5 standard deviations below their TD bilingual peers across the aforementioned tests. Moreover, according to the output of WISC-III assessment (Wechsler, 1992; adapted in Greek by Georgas et al., 2003), both DLD-Mono and DLD-Bi children's verbal abilities were at least 2 standard deviations (*SD*s) below the expected normative mean of chronologically age-matched peers, while their non-verbal scores were within the normal limits for their chronological age (i.e. a non-verbal score of 75 or above; Bloom & Lahey, 1978). None of the children with DLD had received speech and language therapy before inclusion in the study, while all of them attended inclusive classes in schools in which they received literacy skills support by a special education teacher.

All TD bilingual and bilingual children with DLD were second-generation Albanian-Greek children. The two groups were different in terms of family constellations since bilingual children with DLD came from mixed marriages – with one parent being from Greece and the other from Albania – whereas in the case of TD bilingual children both parents came from Albania. The age of acquisition of Greek was also different for the two groups, since bilingual children with DLD were exposed to both languages from birth, whereas for TD bilingual children the mean age of onset to Greek was 4;6 yrs.

Further information about the bilingual children was obtained through a parental questionnaire (Mattheoudakis, Chatzidaki & Maligkoudi, 2014), which was distributed and filled in before the administration of the experimental tasks. The main questions were grouped in two categories: (a) home language history, and

(b) current language use. Home language history referred to exposure to each language from birth up to the age of schooling (i.e. up to the age of six), while current language use referred to language preferences for daily activities (i.e. memorizing phone numbers, calculating, telling the time or watching TV), oral interaction with family members and friends, and the language that the child felt s/he understood or spoke better. For the analysis of the questionnaire data, input in each language was calculated as the proportion of each language used in situations involving the number of people interacting with the child at different stages of development (before 3, between 3 and 6, after 6). For answers that stated that both languages were used to an equal extent, the points were divided between the two languages. Two percentage scores were created, one for Greek and one for Albanian, with 100% and 0% as the maximum and minimum score, respectively. The percentage score for the input that the bilingual child received in Albanian was subsequently subtracted from the input that s/he received in Greek (see Andreou, 2015 for an overview), such that positive scores indicated dominance in Greek, whereas negative scores indicated dominance in Albanian (see Table 2). Analyses of the questionnaire data revealed that TD bilingual children and bilingual children with DLD were different with respect to home language history, $F\,(1, 59) = 105.794$, $p < .001$, $\eta^2 = .80$, since TD-Bi children had significantly higher exposure to Albanian (−18.6) than DLD-Bi children (25.5). The two bilingual groups also differed in the current language use index, $F\,(1, 59) = 4.058$, $p = .049$, $\eta^2 = .26$, since the bilingual group with DLD (31.6) was more Greek-dominant than the TD bilingual group (21.1). This difference may be attributed to the fact that, because children were diagnosed with DLD, their parents wanted them to have societal language input only to improve the children's chances of school integration, which has inevitably led to the attrition of the heritage language. Differences in the age of acquiring Greek may be another reason the two groups differed in language dominance; yet, such differences are reduced significantly when children enter Greek primary educational settings. Both home language history and current language use were entered as covariates in our data analyses.

Participants' socio-economic status (SES) has been derived from maternal education (Ensminger & Fothergill, 2003; Hoff, Laursen & Tardif, 2002) and it was calculated on a 5-point Likert-type scale (adapted from UBILEC; see Unsworth, 2013), with 5 representing the highest educational level attained from compulsory primary education to tertiary education. According to the analysis, there was a significant group effect, $F\,(3, 116) = 54.476$, $p < .001$, $\eta^2 = .23$, which stemmed from the fact that TD-Mono children's mothers had significantly more years of education than the rest of the groups ($p < .001$ for all differences). No significant differences were observed either between DLD-Mono and DLD-Bi ($p > .983$) or between TD-Bi and DLD-Bi children ($p > .998$) (see Table 2). The fact that the mothers of TD-Mono children had much higher education than the rest of the groups may be explained in terms of the socio-economic characteristics of the families of the bilingual Albanian-Greek children and of DLD-Mono children. More specifically, the overwhelming majority of bilingual children came from families of low income and social capital. In fact, most of the Albanian-Greek children's mothers had either dropped out of school or were held back a grade or more before finishing school. On the other hand, the parents of DLD-Mono were of low socio-economic status as well, which explains the parents' preferential use of public diagnostic centers in Greece for their children's assessment, and the fact that none of the children with DLD had received speech and language therapy at a private center before inclusion in the study.

Table 2. Participants' biodata means (and *SDs*)

Group	Home language history (*SD*)	Current language Uuse (*SD*)	SES (*SD*)
TD-Mono (*N* = 30)	–	–	3.7 *(0.5)*
TD-Bi (*N* = 30)	−18.6 *(20.5)*	21.1 *(24.7)*	1.9 *(0.5)*
DLD-Mono (*N* = 30)	–	–	1.8 *(0.9)*
DLD-Bi (*N* = 30)	25.5 *(11.0)*	31.6 *(14.0)*	1.9 *(0.8)*

Note: SES: socio-economic status; TD-Mono: monolingual typically developing children; TD-Bi: bilingual typically developing children; DLD-Mono: monolingual children with Developmental Language Disorder; DLD-Bi: bilingual children with Developmental Language Disorder; *SD*: standard deviation

3.2 General procedure

All four groups of children completed the following tasks in a fixed order: (a) an offline language ability battery consisting of an expressive vocabulary and a sentence repetition task, (b) the story retelling task followed by comprehension questions (MAIN; Gagarina et al., 2012), (c) the online 2-back task, and (d) the online ToM task; as already mentioned, the non-verbal ToM task was only administered to monolingual and bilingual children with DLD. The 2-back and the ToM task were run on a computer using E-Prime software (Schneider, Eschman, & Zuccolotto, 2012). As already mentioned, all tasks were administered in Greek. Children were tested individually at school or in a quiet area of their home by the first and second authors. Participants completed the tasks in two different testing sessions (the 1st included the language ability battery and the retelling task, and the 2nd the 2-back and the ToM task), separated by a minimum duration of 1 week. Data collection took place over a period of six months (September 2017-February 2018).

3.3 Materials and procedure

3.3.1 *Narrative task*

The current study employed the Greek versions of two stories (*Cat* and *Dog*) which were designed for retelling within the COST Action IS0804 (Gagarina et al., 2012; see the Appendix for the input the children listened to in each story). Each story consists of 6 colored pictures that represent the three episodes of the story and involve three main characters (the cat, the butterfly and the boy in the *Cat* story; the dog, the mouse and the boy in the *Dog* story). Each episode includes a goal (e.g. the dog wanted to catch the mouse), an attempt (e.g. the dog leaped forward) and an outcome (e.g. the dog bumped into the tree or the mouse ran away). Half of the children in each experimental group were asked to retell the *Cat* story and the other half the *Dog* story. The story effect in narrative comprehension accuracy was not statistically significant ($p > .10$), so the scores in the two stories were merged.

Each child was first shown three colored envelopes on the computer screen and was asked to pick one of them that included the story. The child listened to the story over headphones while being shown two pictures at a time. After listening to the story, s/he was asked to retell the story to the examiner who has not been listening to the story or looking at the pictures. After retelling the story, the child was asked ten comprehension questions and her/his responses were audiotaped and transcribed separately by the first and second author. Transcripts were then compared word-for-word, with the comparison reaching 98% agreement.

Regarding scoring, the current study fully complied with the guidelines of MAIN (Gagarina et al., 2012). Each correct response received one point (maximum comprehension score: 10 points). Also, the standard MAIN procedure whereby a point can only be given for questions D3, D6, and D9 if the child provides an adequate answer to D2, D5, and D8, respectively, has been followed (see MAIN; Gagarina et al., 2012). Children's responses were scored separately by the first and second author, and the percentage agreement mean (and range) for scoring the answers was 98.9% (97%-100%). Differences between scorings were discussed, changes were made where necessary, and the adjusted scorings were used for the statistical analyses.

As already mentioned, three out of the ten comprehension questions related to the goals of each episode of the story, six questions related to the mental states of the characters and one question assessed children's ability to infer consequences of story events that have taken place, yet, were not explicitly mentioned in the story's aural input. Examples of questions per category and children's answers are provided below:

(1) Question related to the goal of an episode
Why does the dog leap forward?
Child's answer (DLD-Mono, age 6;8): To catch the mouse (correct answer)

(2) Question related to the mental state of a character
Why do you think that the dog feels angry?
Child's answer (DLD-Mono, age 7;1): Because he wants to eat the sausages (wrong answer)

(3) Inference-drawing question
Will the boy be friends with the cat? Why?
Child's answer (DLD-Bi, age 6;6): Yes, because the boy took his ball back (wrong answer)

3.3.2 *Language ability tests*

Before administering the three main experimental tasks of the study (i.e. narrative, 2-back, ToM task), children's language skills were evaluated by measuring their expressive vocabulary and sentence repetition skills. Both language measures have exhibited high sensitivity to language deficits stemming from DLD; expressive vocabulary has been reported to involve DLD-sensitive lexical-semantic processes (Coady, 2013; Theodorou, 2013), while sentence repetition taps into a variety of abilities, including short-term memory, working memory, syntactic and lexical skills (Alloway & Gathercole, 2005), which have been shown to be impaired in children with DLD (Stokes, Wong, Fletcher, & Leonard, 2006; Riches, 2012).

Expressive vocabulary in Greek

Children's expressive vocabulary in Greek was measured through a picture naming test (Vogindroukas, Protopapas, & Sideridis, 2009; the Greek version was adapted from Renfrew, 1997). This task is standardized for 3-to-10-year-old Greek-speaking monolingual children, and consists of 50 black-and-white pictures of objects, which are arranged in order of increasing difficulty and which the child was asked to name. Testing stopped when the child either completed all trials or provided wrong naming (or no response) in five consecutive trials. The highest possible score is 50, with each correct naming response earning one point.

Sentence Repetition Task (SRT)

The Greek SRT employed in the current study was designed within the COST Action IS0804 (Chondrogianni, Andreou, Nerantzini, Varlokosta, & Tsimpli, 2013). The Greek version of the task includes 32 sentences distributed over 8 sets of syntactic structures of varying complexity; namely, Subject-Verb-Object sentences, sentences containing factual and non-factual negation, structures with clitics in clitic left dislocation and clitic doubling contexts, complement clauses, coordinated sentences, adverbial clauses, referential and non-referential object wh-questions, and subject and object relative clauses. In this task, all sentences were grammatical, the eight different structures were matched for length and word frequency, and no fillers were included. There was a 3-item practice session, so that the participants became familiarized with the procedure. During the task, children listened to each sentence via headphones only once and repeated it as accurately as possible. Their responses were audiotaped. Children's sentence repetition performance was assessed for accuracy, i.e. we measured how accurately the child repeated the sentences. If the child's utterance matched the sentence, s/he received 3 points (see Example (4)), whereas in case of one lexical or grammatical substitution, omission or addition s/he received 2 points (see Example (5)). Moreover, if the child made two of the aforementioned errors, s/he received 1 point (see Example (6)), and if the errors were more than three, s/he received zero points. The maximum accuracy score was 96.

(4) Ton kafé ton ípie viastiká o papoús xtes sto kafeneío (correct repetition)
Ton kafé ton
theACC.MASC.SG coffeeACC.MASC.SG himACC.MASC.SG
ípie viastiká o
drinkPAST.ACT.IND.3SG quickly theNOM.MASC.SG
papoús xtes sto kafeneío.
grandpaNOM.MASC.SG yesterday in the shopACC.NEUT.SG
"The grandfather drank the coffee hastily yesterday at the coffee shop."
(Scoring: 3 points)

(5) Ton kafé (omission of the clitic *ton*) ípie viastiká o papoús xtes sto kafeneío
Ton kafé ípie viastiká
theACC.MASC.SG coffeeACC.MASC.SG drinkPAST.ACT.IND.3SG quickly
o papoús xtes sto kafeneío.
theNOM.MASC.SG grandpaNOM.MASC.SG yesterday in the shopACC.NEUT.SG
(Scoring: 2 points)

(6) Ton kafé (omission of the clitic *ton*) ípie (omission of the adverb *viastiká*) o papoús xtes sto kafeneío.
Ton kafé ípie
theACC.MASC.SG coffeeACC.MASC.SG drinkPAST.ACT.IND.3SG
o papoús xtes to kafeneío.
theNOM.MASC.SG grandpaNOM.MASC.SG yesterday in the shopACC.NEUT.SG
(Scoring: 1 point)

3.3.3 *Executive function: 2-back task*

The 2-back task (designed after Pelegrina, Lechuga, García-Madruga, Elosúa, Macizo, Carreiras, Fuentes, & Bajo, 2015; Wild-Wall, Falkenstein, & Gajewski, 2011) required children to monitor the content of a pre-recorded, temporarily presented sequence of digits (2, 5, 7, and 8) at a constant rate of every 4 seconds. The children were asked to remember if the digit they saw on the screen was the same as the one presented two positions back in the sequence; if it was, they were instructed to press a pre-specified key ('J') with their index finger. No responses were required for non-target digits. The task contained a total of 60 trials, out of which 20 were to be responded to (target stimulus trials) and 40 did not trigger a response (non-target trials). The children were familiarized with the task through a practice session of 20 trials, which were not part of the total number of trials. Each trial consisted of a black, 12mm-tall digit that was presented for 500 msecs, followed by a blank page for 2500 msecs, after which the next digit stimulus was presented. An accuracy composite score was computed for each child using the following formula: Composite accuracy = [number of correct hits / number of target stimulus trials (i.e. 20)] minus [number of false hits / number of non-target trials (i.e. 40)]. An example of calculating a child's composite accuracy score is: if correct hits were 15 and wrong hits were 10, then the child's composite accuracy score would be: [15 / 20] – [10 / 40] = 0.75 – 0.25 = 0.50 = 50%. Statistical analyses were conducted on the children's composite accuracy scores. The task was run on E-Prime software (Schneider et al., 2002).

3.3.4 *Online video verification first-order false belief (ToM) task*

Children's non-verbal ToM abilities were tested through an online video verification first-order ToM task adapted from Forgeot d' Arc and Ramous (2010). In this experiment, children were shown a computer screen displaying short video sequences that represented different scenarios having a main agent. Each video sequence consisted of four successive phases. The *beginning phase* introduced the child to the general situation and the main agent, and it was common to all experimental conditions. The *change phase* consisted of five distinct conditions: the 'Mentalistic/Seen change' and the 'Mentalistic/Unseen change' condition that displayed a physical change in the state of the world that was either observed or not observed by the main agent, respectively; the 'Mentalistic/No change' condition wherein no change occurred in the general situation; the 'Mechanistic/Unseen Change' condition, wherein a physical event took place that didn't result from an individual's epistemic state, and the ''Mechanistic/No change' condition, wherein no change occurred, thus, being identical to the 'Mentalistic/No change' condition. The third *suspense phase* was common to all conditions and displayed the main agent coming to the forefront. The *end phase* came in two alternative endings, namely, the 'Mentalistic end' that required from the child to infer the main agent's belief in the given context and thus predict about and track the agent's action, and the 'Mechanistic end' that required from the child to correctly understand and predict the consequence of an event in the physical world without having to determine the main agent's intention or belief. The possible ends in the mentalistic conditions of the task were two videos depicting opposite behavioral outcomes, while the possible ends in the mechanistic conditions were also two videos depicting different physical states.

The task included 10 different scenarios in five conditions, each appearing in two different end versions, coming to a total of 100 video sequence trials (see Forgeot d' Arc & Ramus, 2010 for a detailed description of the plot and the phases across the experimental conditions, as well as pictorial illustrations of the scenarios), and <http://www.lscp.net/persons/forgeot/stim/> for examples and descriptions of each phase across all scenarios). The children completed the task in two sessions spaced less than one week apart. Before each experimental session children underwent a training period with 10 trials to become familiar with the video sequences before the actual experiment began.

After viewing the final video of each scenario, children heard a tone to signal the video sequence was over and viewed a question mark [?] that remained on the center of the screen until the child's response. At this point, children were asked to decide as quickly and accurately as possible whether the end of the scenario was the most appropriate to complete the story they have just seen. Children were instructed to press the green-colored [√] button on a response box if they thought that the way the story ended was plausible, and the red-colored [X] button if they thought the end was inappropriate. Following the answer to the question probe, the next video sequence was initiated by pressing the spacebar. Response times (RT) (i.e. time in msecs from the appearance of the question mark to the child's button-press) and accuracy (%) of judgments was recorded in E-Prime (Schneider et al., 2002). Prior to statistical analyses, subject-by-subject RT on responses which were smaller than 250 msecs and over 2 *SD*s from the mean RT were defined as outliers and were replaced by the mean for each child. This procedure yielded 5.1%, and 4.3% of the data for the DLD-Mono and DLD-Bi group, respectively. These outliers were removed from the data and replaced by the mean for each experimental condition.

For reasons of brevity, analyses will be limited to the accuracy and RT output of (a) the 'Mentalistic Unseen change' condition, which necessitated from children to resolve the interference between their own belief and the belief of the depicted agent who has not witnessed the change in the state of affairs, and (b) the 'Mechanistic Unseen change' condition which did not involve activation of representations of mental states. An accuracy score was computed for each child by calculating the mean percentage of correct decisions and false decisions, and then by subtracting the mean percentage of false decisions from the mean percentage of correct decisions. Reaction times (RT) on correct decisions were also analyzed.

3.4 Analysis plan

All analyses were conducted using the SPSS for Windows software, version 26.0, and the statistical significance level was set at $p < .05$. The study will first report on the group comparisons across the tasks, more specifically, the narrative comprehension task, the two language ability tests, the 2-back and the non-verbal ToM task. A 2 x 2 factorial ANOVA analysis with Bilingualism (monolingual, bilingual) and Disorder (TD, DLD) as the between-subjects variables was run on the data in each task, also controlling for the covariance of age, SES, as well as home language history and current language use for the bilingual groups. Since the non-verbal ToM task was only administered to children with DLD, a repeated measures analysis was run with Change (mentalistic, mechanistic) as the within-subjects variable and group (DLD-Mono, DLD-Bi) as the between-subjects factor separately for the RT and accuracy measure. Finally, a linear regression tested the contribution of age, home language history, current language use, language ability, executive function, and ToM (for children with DLD only) to narrative comprehension. Predictors were entered in a single step and included: chronological age; home language history and current language use, to account for language dominance; expressive vocabulary and sentence repetition scores, to account for language ability; 2-back accuracy scores; and ToM accuracy scores and reactions times in the mentalistic condition (which were available for the two groups with DLD only). The regression was conducted separately by group.

4. Results

4.1 Narrative task

Table 3 provides descriptive statistics for the groups' mean narrative comprehension accuracy scores.[1]

The ANOVA analysis revealed a significant effect for Bilingualism, F (1, 59) = 51.980, $p < .001$, $\eta^2 = .31$, which was due to the fact that bilingual children scored higher than their monolingual peers on narrative comprehension, and a significant Disorder effect, F (1, 59) = 22.590, $p < .001$, $\eta^2 = .17$, which was due to the fact that children with DLD exhibited poorer narrative comprehension than children without DLD. Also, the covariates of chronological age, F (1, 59) = 6.190, $p = .016$, $\eta^2 = .10$, and home language history, F (1, 59) = 5.853, $p = .019$, $\eta^2 = .10$, were found to be

1. Though for the purposes of the present study the groups' narrative comprehension accuracy means were studied, we should note that the pattern of DLD-Mono and DLD-Bi children's narrative comprehension suggests an asymmetry between comprehension questions related to internal states and inferencing, and questions that target the basic episode structure according to Boerma et al.'s (2016) classification. More specifically, while DLD-Mono children performed poorly across-the-board, DLD-Bi children tended to do better in questions referring to the internal states of the characters and requiring inferencing relative to questions relating to the goals of each episode of the stories. These results may provide a database for future work in DLD on potential effects of bilingualism on specific properties of narrative comprehension.

Table 3. Groups' mean accuracy scores (and *SD*s) in narrative comprehension

Task	Groups			
	TD-Mono ($N = 30$)	TD-Bi ($N = 30$)	DLD-Mono ($N = 30$)	DLD-Bi ($N = 30$)
Mean narrative comprehension accuracy score (max. score: 10)	7.6 *(1.0)*	8.2 *(0.7)*	5.3 *(1.3)*	8.2 *(1.9)*

Note: TD-Mono: monolingual typically developing children; TD-Bi: bilingual typically developing children; DLD-Mono: monolingual children with Developmental Language Disorder; DLD-Bi: bilingual children with Developmental Language Disorder; *SD*: standard deviation.

significant. The analysis has also revealed a significant interaction between Bilingualism and Disorder, F (1, 59) = 24.350, $p < .001$, $\eta^2 = .17$. To unpack the significant interaction, independent samples t-tests were conducted. The tests revealed that TD-Bi and DLD-Bi children had significantly higher narrative comprehension accuracy scores than their TD-Mono and DLD-Mono peers (t (58) = 2.386, $p = .024$, and t (58) = 6.759, $p < .001$, respectively). Also, DLD-Mono children had significantly lower scores than TD-Mono children, t (58) = 6.750, $p < .001$, while no difference was observed between bilingual children with and without DLD, and t (58) = .081, $p = .936$.

4.2 Language Ability tests

Table 4 provides descriptive statistics of the children's mean accuracy scores in expressive vocabulary and sentence repetition. The analysis on the expressive vocabulary data revealed a significant effect for Disorder, F (1, 59) = 7.694, $p = .008$, $\eta^2 = .12$, which stemmed from the fact that children with DLD had lower vocabulary scores than their TD peers, as well as a significant Bilingualism effect, F (1, 59) = 20.716, $p < .001$, $\eta^2 = .15$, which stemmed from the fact that bilingual children scored lower than their monolingual peers on expressive vocabulary. The covariates of age, F (1, 59) = 5.623, $p = .021$, $\eta^2 = .09$, and current language use, F (1, 59) = 9.448, $p = .003$, $\eta^2 = .15$, were found to be significant, while the interaction between Bilingualism and Disorder was not found to be significant, F (1, 59) = .813, $p = .369$, $\eta^2 = .01$. Interestingly, though DLD-Bi children were more Greek-dominant than their TD-Bi peers (see Table 2), this difference did not seem to be reflected in children's expressive vocabulary scores, since the TD-Bi group had higher accuracy scores compared to bilingual children with DLD.

The analysis on the groups' sentence repetition scores revealed a significant effect for Disorder, F (1, 59) = 16.323, $p = .008$, $\eta^2 = .23$, which stemmed from the fact that children with DLD has lower sentence repetition scores than their TD peers, a significant effect for home language history, F (1, 59) = 5.105, $p = .028$, $\eta^2 = .09$, and a significant interaction between Bilingualism and Disorder, F (1, 59) = 7.026,

$p = .009$, $\eta^2 = .06$. The independent samples t-tests revealed that DLD-Mono children outperformed their DLD-Bi peers in sentence repetition, $t\ (58) = 4.644$, $p < .001$; the difference between the two TD groups was not found to be significant, $t\ (58) = 1.485$, $p = .148$. Also, DLD-Mono children had significantly lower scores than TD-Mono children, $t\ (58) = 5.961$, $p < .001$, and DLD-Bi children performed significantly poorer than TD-Bi children, $t\ (58) = 2.598$, $p = .015$.

Table 4. Expressive vocabulary and sentence repetition scores by Group

Group	Mean expressive vocabulary score (max. score: 50) (*SD*)	Mean sentence repetition score (max. score: 96) (*SD*)
TD-Mono (*N* = 30)	33.6 *(3.2)*	77.5 *(9.1)*
TD-Bi (*N* = 30)	30.6 *(3.2)*	71.4 *(6.7)*
DLD-Mono (*N* = 30)	30.7 *(5.5)*	61.9 *(15.0)*
DLD-Bi (*N* = 30)	26.7 *(3.0)*	55.5 *(17.7)*

Note: TD-Mono: monolingual typically developing children; TD-Bi: bilingual typically developing children; DLD-Mono: monolingual children with Developmental Language Disorder; DLD-Bi: bilingual children with Developmental Language Disorder; *SD*: standard deviation

4.3 Executive function: 2-back task

Table 5 provides descriptive statistics of the children's composite accuracy scores in the 2-back task.

Table 5. Groups' mean composite accuracy scores (%) in the 2-back task

Group	2-back accuracy (%) (*SD*)
TD-Mono (*N* = 30)	30.6 *(13.6)*
TD-Bi (*N* = 30)	55.8 *(21.5)*
DLD-Mono (*N* = 30)	10.2 *(9.5)*
DLD-Bi (*N* = 30)	6.5 *(14.0)*

Note: TD-Mono: monolingual typically developing children; TD-Bi: bilingual typically developing children; DLD-Mono: monolingual children with Developmental Language Disorder; DLD-Bi: bilingual children with Developmental Language Disorder; *SD*: standard deviation

The analysis revealed a significant effect for Disorder, F (1, 59) = 29.453, p < .001, η^2 = .35, which stemmed from the fact that children with DLD had lower accuracy scores than their TD peers, a significant effect for Bilingualism, F (1, 59) = 14.677, p < .001, η^2 = .11, which stemmed from the fact that bilingual children had higher accuracy scores than their monolingual peers, a significant effect for the covariate of age, F (1, 59) = 5.883, p = .019, η^2 = .10, and a significant two-way interaction between Bilingualism and Disorder, F (1, 59) = 26.684, p < .001, η^2 = .19. The independent samples t-tests revealed that TD-Bi children outperformed their TD-Mono peers in 2-back accuracy, t (58) = 5.576, p < .001. However, the difference between the two groups with DLD was not found to be significant, t (8) = .477, p = .637. Also, both DLD-Mono and DLD-Bi children had significantly lower scores than TD-Mono and TD-Bi children (t (58) = 7.405, p < .001, and t (58) = 10.184, p = .015, respectively).

4.4 Online video verification first-order false belief task

Table 6 provides descriptive statistics for DLD-mono and DLD-bi children's accuracy and RT performance in the 'Mechanistic Unseen change' and 'Mentalistic Unseen change' condition of the online false belief task. Prior to statistical analyses, subject-by-subject RT on judgments which were smaller than 250 msecs and over 2 SDs from the mean RT for each condition (Mechanistic, Mentalistic) were considered outliers and replaced by the mean.

Table 6. DLD groups' Mean accuracy scores (%) (and *SD*s) and RT on correct decisions in the 'Mechanistic Unseen change' and 'Mentalistic Unseen change' conditions of the online video verification first-order false belief task

Group	Mechanistic Unseen change		Mentalistic Unseen change	
	Accuracy (%) (*SD*)	RT (*SD*)	Accuracy (%) (*SD*)	RT (*SD*)
DLD-Mono (N = 30)	75.1 (*10.2*)	3154 (*475*)	52.9 (*9.7*)	4680 (*739*)
DLD-Bi (N = 30)	73.9 (*10.5*)	2029 (*705*)	63.7 (*13.1*)	2789 (*573*)

Note: DLD-Mono: monolingual children with Developmental Language Disorder; DLD-Bi: bilingual children with Developmental Language Disorder; *SD:* standard deviation

We first report the results of the RT measure. The repeated measures analysis revealed a significant group effect, F (1, 57) = 27.461, p < .001, η^2 = .33, which stemmed from the fact that DLD-Bi children were significantly faster than their DLD-Mono peers, and a significant interaction between group and Change, F (1, 57) = 4.336,

$p = .042$, $\eta^2 = .07$. The effect for age was not found to be significant, $F(1, 57) = 1.818$, $p = .183$, $\eta^2 = .03$. Subsequent paired samples *t*-tests revealed that mentalistic trials were responded significantly faster than mechanistic trials for both DLD-Mono and DLD-Bi groups ($t(29) = 5.061$, $p < .001$, and $t(1, 29) = 2.970$, $p = .006$, respectively). Moreover, independent samples *t*-tests revealed that DLD-Mono children were significantly slower than their DLD-Bi peers in both mechanistic, $t(58) = 4.549$, $p < .001$, and mentalistic trials, $t(58) = 3.267$, $p = .003$.

Regarding accuracy, the same analysis revealed non-significant effects for either group, $F(1, 59) = 1.781$, $p = .187$, $\eta^2 = .03$, or age, $F(1, 59) = 2.861$, $p = .096$, $\eta^2 = .04$; yet, there was a significant interaction between group and Change, $F(1, 57) = 3.622$, $p = .048$, $\eta^2 = .06$. The paired samples t-tests for each group revealed that mechanistic trials were responded more accurately than mentalistic trials only for the DLD-Mono group, $t(1, 29) = 4.727$, $p < .001$; the difference in accuracy between mentalistic and mechanistic trials for the DLD-Bi group was not found to be significant, $t(1, 29) = 1.610$, $p = .118$. Also, independent samples *t*-tests revealed that DLD-Mono children were significantly less accurate than their DLD-Bi peers in mentalistic trials only, $t(58) = 1.792$, $p < .001$; the two groups did not differ in mechanistic trials, $t(58) = .278$, $p = .385$.

4.5 Narrative comprehension, age, language dominance and independent language ability, executive function and ToM assessments

The linear regression analysis tested the links of narrative comprehension with age, home language history, current language use, language ability, executive function, and ToM for each group (see Table 7). For the TD-Mono group, the overall model was significant, $F(4, 29) = 113.759$, $p < .001$; $R^2 = .94$, $p < .001$. In this model, two predictors accounted for significant independent variance: 2-back accuracy, $\beta = .613$, $p < .001$; and sentence repetition, $\beta = .331$, $p = .002$. For the TD-Bi group, the overall model was also significant, $F(6, 29) = 77.975$, $p < .001$; $R^2 = .91$, $p < .001$, while variance in narrative comprehension was significantly accounted for by two predictors: chronological age, $\beta = .273$, $p = .05$; and 2-back accuracy, $\beta = .719$, $p < .001$. For DLD-Mono children, the overall model was significant, $F(6, 29) = 10.618$, $p < .001$; $R^2 = .67$, $p < .001$, yet, variance in narrative comprehension was accounted for by children's chronological age only, $\beta = .868$, $p < .001$. Finally, for DLD-Bi children, the overall model was significant, $F(8, 29) = 161.332$, $p < .001$; $R^2 = .97$, $p < .001$. In this model, five predictors accounted for significant independent variance: chronological age, $\beta = .274$, $p = .05$; home language history, $\beta = .348$, $p = .047$; current language use, $\beta = .561$, $p = .002$; accuracy in mentalistic trials, $\beta = .444$, $p = .005$, and RT in mentalistic trials, $\beta = -.275$, $p = .044$. Table 7 presents the entire models with information on both significant and non-significant predictors.

Table 7. Results of the linear regression analysis on the narrative comprehension performance of each group

Group	Independent predictors							
	Age	Home language history	Current language use	Expressive vocabulary	Sentence repetition	2-back accuracy	ToM/accuracy	ToM/RT
TD-Mono ($N = 30$)	$\beta = .064$, $p = .359$	–	–	$\beta = .041$, $p = .389$	$\beta = .331$, $p = .002$	$\beta = .613$, $p < .001$	–	–
TD-Bi ($N = 30$)	$\beta = .273$, $p = .05$	$\beta = .355$, $p = .113$	$\beta = .089$, $p = .689$	$\beta = .035$, $p = .602$	$\beta = .135$, $p = .165$	$\beta = .719$, $p < .001$	–	–
DLD-Mono ($N = 30$)	$\beta = .868$, $p < .001$	–	–	$\beta = .126$, $p = .396$	$\beta = .076$, $p = .566$	$\beta = .169$, $p = .271$	$\beta = .073$, $p = .654$	$\beta = -.072$, $p = .559$
DLD-Bi ($N = 30$)	$\beta = .274$, $p = .05$	$\beta = .348$, $p = .047$	$\beta = .561$, $p = .002$	$\beta = .170$, $p = .311$	$\beta = .161$, $p = .282$	$\beta = .033$, $p = .314$	$\beta = .444$, $p = .005$	$\beta = -.275$, $p = .044$

Note: TD-Mono: monolingual typically developing children; TD-Bi: bilingual typically developing children; DLD-Mono: monolingual children with Developmental Language Disorder; DLD-Bi: bilingual children with Developmental Language Disorder; ToM: Theory of Mind

5. Discussion

The current study compared the narrative comprehension skills after listening and retell in four groups of children: two groups of 6 to 8 year old monolingual and bilingual children with DLD and two groups of age-matched TD monolingual and bilingual children. The study was carried out with the use of the Greek versions of two stories (*Cat* and *Dog*) which have been designed for retelling in the Multilingual Assessment Instrument for Narratives (Gagarina et al., 2012). The study also explored group performance in tasks tapping into language, executive function and ToM abilities to determine whether narrative comprehension performance relates to any of these skills. First, the results revealed a significant bilingualism effect on the narrative comprehension skills of children with and without DLD; both TD-Bi and DLD-Bi groups exhibited higher accuracy scores than their monolingual peers in comprehension questions. Regarding the rest of the tasks, bilingualism was found to contribute to better mentalizing abilities in children with DLD. However, given that the non-verbal ToM task was not administered to the TD groups of the study, we cannot draw conclusions about the role of mentalizing skills in narrative comprehension in TD monolingual and bilingual children. Finally, regression analyses, which were conducted separately by group, indicated that performance in the narrative comprehension task was associated with the executive function measure for the TD groups, and with ToM and language dominance for the DLD-Bi group, while the comprehension performance of DLD-Mono children was only associated with chronological age.

The first aim of the study was to investigate the effect of disorder, bilingualism, and their combined effect on the experimental groups' narrative comprehension. Narrative comprehension was not restricted to the comprehension of picture stories only, but it was tested after listening and retell. In both bilingual groups with and without DLD, children achieved higher narrative comprehension scores than their monolingual peers. This finding comes into contrast with Boerma and colleagues' (2016) study which failed to detect any bilingualism effect on DLD children's narrative comprehension skills. The difference in findings may be attributed to the different profiles of the bilingual children with DLD that have participated in the two studies. First, children in Boerma et al.'s (2016) study were younger (5 to 6 year old) than the children in the current study (6 to 8 year old), which may have affected their comprehension performance. Crucially, according to the regression analyses of the present study, narrative comprehension scores for all experimental groups except for TD monolingual children were found to increase with age, which suggests that narrative comprehension in DLD may be subject to maturational-dependent processes that become progressively active with increasing age. Besides age, bilingual children with DLD in the current study may have also benefited from their dominance in L2/Greek which was the language in which the narrative data were elicited. Such benefit seems to be verified by both the covariance analyses that show that narrative comprehension scores in bilingual children increased with higher exposure to L2/Greek and the regression analyses according to which dominance in L2/Greek accounted for most of DLD-Bi children's variance in narrative comprehension.

In line with relevant literature on the role of executive functions in TD children's narrative comprehension (Kim, 2016), updating skills seemed to contribute to both TD-Mono and TD-Bi children's narrative comprehension performance. In fact, the strongest benefit from updating was observed for TD-Bi children who outperformed their monolingual peers and children with DLD in the 2-back task. On the other hand, narrative comprehension did not seem to be affected by the executive function measure in children with DLD. In fact, both monolingual and bilingual children with DLD performed equally low in the 2-back task with a mean accuracy score that was slightly above zero level (≤ 10%). The specific task may have imposed too high demands on DLD children's non-verbal working memory to allow links to emerge with the narrative comprehension. It is argued that *n*-back tasks are, in general, more resource demanding than tasks tapping into inhibition or/and attention, due to requiring the simultaneous planning, co-ordination and monitoring of the subject's responses (Wild-Wall et al., 2011). Specifically, children had to simultaneously monitor a series of digits appearing on the screen for a limited period of time, update in their working memory a continuous stream of information to integrate recently presented digits, attend to specific digits and inhibit attention to interfering stimuli. These cognitive operations can only be fulfilled if meticulously coordinated. All these factors might have aggravated the difficulty of the 2-back task which proved to be particularly demanding for both monolingual and bilingual children with DLD. The asymmetry in bilingualism effects in TD children and children with DLD appears to be in line with Engel de Abreu and colleagues' (2014) study, in which TD bilingual children outperformed their monolingual peers on an interference suppression task, yet, bilingual children with DLD performed significantly worse than TD bilinguals. Though we cannot pin down the exact deficit which might have been responsible for both DLD groups' low performance in the 2-back task, the findings speak in favor of the occurrence of a central attentional bottleneck in children with DLD which potentially led to processing deficits that could not be compensated for by bilingualism.

The third aim of the study was to evaluate DLD children's performance in a non-verbal ToM task. The DLD-Bi group outperformed their monolingual peers in both the accuracy and RT measure of the task. Specifically, DLD-Bi children were significantly more accurate than their DLD-Mono peers in the mentalistic (vs. mechanistic) trials of the task, which implies that bilingual children performed more efficiently than monolinguals only in the experimental condition that called upon the children's mental state attribution skills. Besides accuracy, the two groups also differed in their response times in mentalistic trials, with DLD-Bi children being considerably faster than DLD-Mono children, for whom reasoning about other individuals' beliefs was probably more effortful. Theory of Mind is a vulnerable skill in children with language impairment, as evinced in studies using verbal ToM tasks (Farrant, 2015; Nilsson & de López, 2016 for a review; Spanoudis, 2016; Tsimpli et al., 2017). Importantly, DLD children's ToM skills in the present study were evaluated through an online non-verbal measure so as to overcome previous methodological shortcomings stemming from high correlations found between verbal false belief reasoning tasks and general language ability (Astington & Jenkins, 1999; de Villiers & de Villiers, 2000). A limitation of the current study stems from the fact that the non-verbal ToM task was only administered to children with DLD with no baseline data from TD children. Though this limitation prevents us from drawing any strong conclusions about the extent to which false belief attribution abilities were affected by bilingualism in children with DLD, the current finding speaks in favor of a boosting effect on children's mentalizing abilities.

The final aim of the study was to examine the contribution of language and higher-order cognitive skills to each experimental group's narrative comprehension performance. Previous research (Tsimpli et al., 2016, 2017) exploring the relationship between language ability and narrative performance in bilingual children with DLD has exclusively focused on language production. In this study, we directly assessed the influence of language ability, executive functions and ToM, as well as language dominance, on the narrative comprehension of monolingual and bilingual children with and without DLD.

Regarding TD children, the regression analyses indicated that performance in the narrative comprehension task was associated with both sentence repetition and updating skills for TD-Mono children, while performance for TD-Bi children was mostly associated with their non-verbal updating skills. We observed that TD-Bi and TD-Mono children had a similar performance in their language abilities (see Table 4). This result may have been driven by the fact that TD-Bi children attended Greek monolingual educational settings with no literacy support in their mother tongue. In Table 2 that presents participants' biodata means, we observe that, although in past activities the TD-Bi group appeared to be as Albanian-dominant, the picture is different when it comes to current activities where the same group appears to be as Greek-dominant. In other words, literacy in Greek and the increase of input in Greek may be responsible for the fact that TD-Bi children exhibited the same performance with their monolingual peers in grammatical abilities. If we want to further explain TD groups' performance in the narrative comprehension task we may say that since both TD-Mono and TD-Bi children have reached a high threshold in vocabulary and grammatical abilities, it is the individual differences of each group in executive functioning that seemed to lead their narrative comprehension performance.

Regarding children with DLD, ToM proved to be important for narrative comprehension in the DLD-Bi group, while DLD-Mono children's performance was only predicted by their chronological age. Inspection of the data patterns shows that both groups with DLD tended to draw on fewer executive and linguistic resources relative to their TD peers, which further suggests that these abilities had little impact on DLD children's narrative comprehension. One possible reason for the limited role of language ability in DLD-Mono children's narrative comprehension is that the lexical and grammatical requirements of narrative comprehension were satisfied by the children's language proficiency level. The assumption that the lexical and grammatical requirements of narrative comprehension were below the threshold of DLD children's language skills receives support from the finding that DLD-Bi children's dominance in L2/Greek had a strong impact on their efficiency of narrative comprehension.

Though the two groups with DLD did not differ in expressive vocabulary and in 2-back accuracy, they had a disproportionately larger difference in performance in the non-verbal ToM task in both the accuracy and the RT measure. The fact that ToM was a significant predictor of DLD-Bi children's narrative comprehension performance suggests that differences between the two groups with DLD in this dimension, i.e. the domain of meta-representational reasoning skills, contributed to the group differences in comprehension. Of course, though accuracy scores and RT in the non-verbal ToM task contributed independent variance to DLD-Bi children's narrative comprehension, i.e. narrative comprehension was more efficient as accuracy in the ToM task increased and RT on their responses were faster, the data do not conclusively indicate that high ToM skills led directly to better narrative comprehension. One consideration is that bilingual children with DLD could perform the task using an alternative strategy, e.g. executive function sub-processes that have not been affected by language impairment, such as visuospatial working memory (Engel de Abreu et al., 2014). Future studies must explore the relationship between narrative comprehension and other executive functions in both monolingual and bilingual children with DLD, so as to determine their contribution to children's narrative performance.

6. Conclusions

The present study sheds light on issues relating to bilingualism effects in the narrative comprehension skills of 6 to 8 year old Albanian-Greek children with and without DLD using the Greek versions of the stories (*Cat* and *Dog*) designed for retelling in the Multilingual Assessment Instrument for Narratives (Gagarina et al., 2012). The study has also explored possible relations between children's narrative comprehension performance and their language ability, executive function and ToM skills. Bilingual children with DLD performed better than their monolingual peers on narrative comprehension, though no between-group differences were found for expressive vocabulary and executive function abilities. The findings from the non-verbal ToM task, moreover, suggest that bilingualism may help boost DLD children's false belief attribution abilities. Such an effect was reflected in faster response times and higher accuracy for bilingual children with DLD as compared to monolingual children with DLD in trials of the task that tapped into children's meta-representational reasoning skills. Results also indicate that narrative comprehension in children with DLD was not associated with executive function skills, in contrast to their TD peers. Instead, narrative comprehension in bilingual children with DLD was influenced by ToM and proficiency in L2/Greek. The overall evidence highlights advantages for bilingual children with DLD in narrative comprehension and ToM skills, also suggesting that ToM skills afford bilingual children with DLD the opportunity for more efficient meaning-integration processes in narrative comprehension as compared to their monolingual peers with DLD.

Acknowledgements

We would like to thank the children and their parents for their participation in our study.

References

Alloway, T. P., & Gathercole, S. E. (2005). The role of sentence recall in reading and language skills of children with learning difficulties. *Learning and Individual Differences*, 15(4), 271–282. https://doi.org/10.1016/j.lindif.2005.05.001

Altman, C., Armon-Lotem, S., Fichman, S., & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English – Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics*, 37(1), 165–193. https://doi.org/10.1017/S0142716415000466

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. Washington, DC. https://doi.org/10.1176/appi.books.9780890425596

Andreou, M. (2015). The effects of bilingualism on verbal and non verbal cognition: The micro- and macro-structure of narratives in the weak and the dominant language of the bilingual child (Unpublished doctoral dissertation). Aristotle University of Thessaloniki.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, 35(5), 1311–1320. https://doi.org/10.1037/0012-1649.35.5.1311

Atkinson, L., Slade, L., Powell, D., & Levy, J. P. (2017). Theory of mind in emerging reading comprehension: A longitudinal study of early indirect and direct effects. *Journal of Experimental Child Psychology*, 164, 225–238. https://doi.org/10.1016/j.jecp.2017.04.007

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, 19(1), 19–23. https://doi.org/10.1177/0963721409358571

Bialystok, E., & Senman, S. (2004). Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development*, 75(2), 562–579. https://doi.org/10.1111/j.1467-8624.2004.00693.x

Blom, E., & Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure? *Research in Developmental Disabilities*, 55, 301–311. https://doi.org/10.1016/j.ridd.2016.05.001

Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. New York, NY: Wiley & Sons.

Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: Implications for clinical practice. *Journal of Language and Communication Disorders*, 5(6), 626–638. https://doi.org/10.1111/1460-6984.12234

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics*, 37(1), 19–48. https://doi.org/10.1017/S0142716415000405

Chondrogianni, V., Andreou, M., Nerantzini, M., Varlokosta, S., & Tsimpli, I. M. (2013). *The Greek Sentence Repetition Task*. COST Action IS0804.

Chrysochoou, E., & Bablekou, Z. (2011). Phonological loop and central executive contributions to oral comprehension skills of 5.5 to 7.5 years old children. *Applied Cognitive Psychology*, 25(4), 576–583. https://doi.org/10.1002/acp.1723

Coady, J. A. (2013). Rapid naming by children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 56(2), 604–617. https://doi.org/10.1044/1092-4388(2012/10-0144)

Daneman, M., & Blennerhassett, A. (1984). How to assess the listening comprehension skills of prereaders. *Journal of Educational Psychology*, 76(6), 1372–1381. https://doi.org/10.1037/0022-0663.76.6.1372

de Villiers, P. A. (1991). English literacy development in deaf children: Directions for research and intervention. In J. Miller (Ed.), *Research in child language disorders: A decade of progress* (pp. 349–378). Austin TX: ProEd.

de Villiers, J. G., & de Villiers, P. A. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 189–226). Hove: Psychology Press.

Dennett, D. C. (Eds.). (1978). *Brainstorms: Philosophical essays on mind and psychology*. Cambridge, MA: The MIT Press.

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: An investigation of their narratives and memory. *International Journal of Communication and Language Disorders*, 43(2), 201–218. https://doi.org/10.1080/13682820701366147

Engel de Abreu, P. M. J., Cruz-Santos, A., & Puglisi, M. L. (2014). Specific language impairment in language-minority children from low-income families. *Journal of Language & Communication Disorders*, 49(6), 736–747. https://doi.org/10.1111/1460-6984.12107

Ensminger, M. E., & Fothergill, K. E. (2003). A decade of measuring SES: what it tells us and where to go from here. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting and child development* (pp. 13–27). Mahwah, NJ: Lawrence Erlbaum Associates.

Farrant, B. (2015). Specific language impairment and perspective taking: Delayed development of theory of mind, visual and emotional perspective taking. *Journal of Childhood and Developmental Disorders*, 1(8).

Florit, E., Roch, M., Altoe, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology*, 27(4), 935–951. https://doi.org/10.1348/026151008X397189

Florit, E., Roch, M., & Levorato, M. C. (2011). Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. *Discourse Processes*, 48(2), 119–138. https://doi.org/10.1080/0163853X.2010.494244

Florit, E., Roch, M., & Levorato, M. C. (2013). The relationship between listening comprehension of text and sentences in preschoolers: Specific or mediated by lower- and higher-level components? *Applied Psycholinguistics*, 34(2), 395–415. https://doi.org/10.1017/S0142716411000749

Ford, J. A., & Milosky, L. M. (2003). Inferring emotional reactions in social situations: differences in children with language impairment. *Journal of Speech, Language, and Hearing Research*, 46(1), 21–30. https://doi.org/10.1044/1092-4388(2003/002)

Forgeot d' Arc, B., & Ramus, F. (2010). Belief attribution despite verbal interference. *The Quarterly Journal of Experimental Psychology*, 64(5), 975–990. https://doi.org/10.1080/17470218.2010.524413

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics, 56*. Berlin: Leibniz Zentrum für Allgemeine Sprachwissenschaft.

Georgas, J., Paraskevopoulos, I. N., Besevegis, E., Giannitsas, N., & Mylonas, K. (2003). Greece. In J. Georgas, L. G. Weiss, F. J. R. van de Vijver, & D. H. Saklofske (Eds.), *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (pp. 199–214). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-012280055-9/50015-1

Gillott, A., Furniss, F., & Walter, A. (2004). Theory of mind ability in children with specific language impairment. *Child Language Teaching and Therapy*, 20(1), 1–11. https://doi.org/10.1191/0265659004ct260oa

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition*, 6(1), 1–15. https://doi.org/10.1017/S1366728903001007

Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M. Bornstein (Eds.), *Handbook of parenting, Vol. 2: Biology and ecology of parenting*, (pp. 231–252). Mahwah, NJ: Lawrence Erlbaum Associates.

Karasinski, C., & Ellis Weismer, S. (2010). Comprehension of inferences in discourse processing by adolescents with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 53(5), 1268–1279. https://doi.org/10.1044/1092-4388(2009/09-0006)

Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension: A critical review. *Topics in Language Disorders*, 33(3), 208–233. https://doi.org/10.1097/TLD.0b013e31829d623e

Kim, Y. S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology*, 141, 101–120. https://doi.org/10.1016/j.jecp.2015.08.003

Laloi, A., de Jong, J., & Baker, A. (2017). Can executive functioning contribute to the diagnosis of SLI in bilingual children? *Linguistic Approaches to Bilingualism*, 7(3), 431–459. https://doi.org/10.1075/lab.15020.lal

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: The MIT Press.

Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4- to 6-year-old children. *Reading Research Quarterly*, 47(3), 259–282. https://doi.org/10.1002/rrq.020

Lukács, Á., Ladányi, E., Fazekas, K., & Kemény, F. (2016). Executive functions and the contribution of short-term memory span in children with Specific Language Impairment. *Neuropsychology*, 30(3), 296–303. https://doi.org/10.1037/neu0000232

Lynch, J. S., & van den Broek, P. (2007). Understanding the glue of narrative structure: children's on- and off-line inferences about characters' goals. *Cognitive Development*, 22(3), 323–340. https://doi.org/10.1016/j.cogdev.2007.02.002

Marton, K. (2008). Visuo-spatial processing and executive functions in children with specific language impairment. *International Journal of Language & Communication Disorders*, 43(2), 181–200. https://doi.org/10.1080/16066350701340719

Marton, K., Campanelli, L., Eichorn, N., Scheuer, J., & Yoon, J. (2014). Information processing and proactive interference in children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 57(1), 106–119. https://doi.org/10.1044/1092-4388(2013/12-0306)

Mattheoudakis, M., Chatzidaki, A., & Maligkoudi, C. (2014). Bilingual education & types of bilingualism. Paper presented at the Workshop "Language Knowledge and Development in native and non-native speakers", Aristotle University of Thessaloniki, Greece.

Maviş, I., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. *Applied Psycholinguistics*, 37(1), 69–89. https://doi.org/10.1017/S0142716415000429

Miller, C. A. (2001). False belief understanding in children with specific language impairment. *Journal of Communication Disorders*, 34(1–2), 73–86.

Nilsson, K. K., & de López, K. J. (2016). Theory of Mind in children with specific language impairment: A systematic review and meta-analysis. *Child Development*, 87(1), 143–153. https://doi.org/10.1111/cdev.12462

Oakhill, J. V., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, 18(7–9), 657–686. https://doi.org/10.1007/s11145-005-3355-z

Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 1–25. https://doi.org/10.1080/13670050.2018.1434124

Pelegrina, S., Lechuga, M. T., García-Madruga, J. A., Elosúa, M. R., Macizo, P., Carreiras, M., Fuentes, L. J., & Bajo, M. T. (2015). Normative data on the n-back task for children and young adolescents. *Frontiers in Psychology*, 6, 1544. https://doi.org/10.3389/fpsyg.2015.01544

Peristeri, E., Baldimtsi, E., Tsimpli, I. M., & Durrleman, S. (2019). Bilingualism effects in children with Developmental Language Disorder: Metalinguistic awareness, executive functions, and false-belief reasoning. In M. M. Brown, & B. Dailey (Eds.), *BUCLD 43: Proceedings of the 43rd annual Boston University Conference on Language Development* (pp. 549–560). Boston, MA: Cascadilla Press.

Renfrew, C. (1997). *Word Finding Vocabulary Test* (The Renfrew Language Scales). Oxford: Winslow Press.

Rezzonico, S., Chen, X., Cleave, P., Greenberg, J., Hipfner-Boucher, K., Johnson, C., & Girolametto, L. (2015). Oral narratives in monolingual and bilingual preschoolers with SLI. *International Journal of Language & Communication Disorders*, 50(6), 830–841. https://doi.org/10.1111/1460-6984.12179

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. *International Journal of Language & Communication Disorders*, 47(5), 499–510. https://doi.org/10.1111/j.1460-6984.2012.00158.x

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Spanoudis, G. (2016). Theory of mind and specific language impairment in school-age children. *Journal of Communication Disorders*, 61, 83–96. https://doi.org/10.1016/j.jcomdis.2016.04.003

Stavrakaki, S., & Tsimpli, I. M. (2000). Diagnostiko test glossikis noimosinis gia pedia prosholikis ke sholikis ilikias: stathmisi, statistiki analisi ke psihometrikes idiotites. [Diagnostic Verbal IQ Test for Greek preschool and school age children: standardization, statistical analysis, and psychometric properties]. In M. Glykas, & G. Kalomiris (Eds.), *Proceedings of the 8th conference of the Pan-Hellenic Association of Speech and Language Therapists* (pp. 95–106). Athens: Ellinika Grammata.

Stokes, S. F., Wong, A. M. Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research*, 49(2), 219–236. https://doi.org/10.1044/1092-4388(2006/019)

Strasser, K., & del Rio, F. (2014). The role of comprehension monitoring, Theory of Mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Reading Research Quarterly*, 49(2), 169–187. https://doi.org/10.1002/rrq.68

Theodorou, E. (2013). Specific Language Impairment in CYPRIOT Greek: Diagnostic and experimental investigations (Unpublished doctoral dissertation). University of Cyprus.

Tsimpli, I. M., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with Specific Language Impairment. *Applied Psycholinguistics*, 37(1), 195–216. https://doi.org/10.1017/S0142716415000478

Tsimpli, I. M., Peristeri, E., & Andreou, M. (2017). Clitic production in monolingual and bilingual children with Specific Language Impairment: Evidence from elicitation and narrative tasks. *Linguistic Approaches to Bilingualism*, 7(3–4), 394–430. https://doi.org/10.1075/lab.15025.tsi

Unsworth, S. (2013). *UBILEC: Utrecht bilingual language exposure calculator* (Unpublished manuscript).
Vissers, C., Koolen, S., Hermans, D., Scheper, A., & Knoors, H. (2015). Executive functioning in preschoolers with specific language impairment. *Frontiers in Psychology*, 6, 1574. https://doi.org/10.3389/fpsyg.2015.01574
Vogindroukas, I., Protopapas, A., & Sideridis, G. (2009). *Expressive vocabulary assessment* (Greek version of Renfrew Word Finding Vocabulary Test). Chania, Crete: Glafki.
Was, C. A., & Woltz, D. J. (2007). Reexamining the relationship between working memory and comprehension: The role of available long-term memory. *Journal of Memory and Language*, 56(1), 86–102. https://doi.org/10.1016/j.jml.2006.07.008
Wechsler, D. (1992). *WISC-III: Wechsler Intelligence Scale for Children – Third Edition: Manual (Australian adaptation)*. San Antonio, TX: Psychological Corporation.
Wild-Wall, N., Falkenstein, M., & Gajewski, P. D. (2011). Age-related differences in working memory performance in a 2-back task. *Frontiers in Psychology*, 2, 186. https://doi.org/10.3389/fpsyg.2011.00186
Ziatas, K., Durkin, K., & Pratt, C. (1998). Belief term development in children with autism, Asperger syndrome, specific language impairment, and normal development: Links to theory of mind development. *Journal of Child Psychology and Psychiatry*, 39(5), 755–763. https://doi.org/10.1111/1469-7610.00374

Appendix

<Cat story>

Greek text:
Μια μέρα ήταν μια παιχνιδιάρα γάτα που είδε ότι μια πεταλούδα καθόταν πάνω σε ένα θάμνο. Πήδηξε ψηλά γιατί ήθελε να τη πιάσει. Στο μεταξύ, ένα χαρωπό αγόρι γυρνούσε από το ψάρεμα κρατώντας ένα κουβά και μια μπάλα στα χέρια του. Είδε ότι η γάτα κάτι κυνηγούσε. Το αγόρι είπε: «Γάτα, γάτα, τι κυνηγάς;» Όμως η πεταλούδα πέταξε μακριά και η γάτα έπεσε πάνω στο θάμνο. Το αγόρι τρόμαξε και του έπεσε η μπάλα από το χέρι. Αλλά ήθελε την μπάλα του και αποφάσισε να την πιάσει. Στο μεταξύ, η γάτα πρόσεξε το κουβά του αγοριού και σκέφτηκε: Τι να άφησε το αγόρι στο κουβά; Το αγόρι άρχισε να τραβάει τη μπάλα του έξω από το νερό. Την ίδια ώρα, η γάτα έφτανε το ψάρι που άφησε το αγόρι και σκεφτόταν: «Αυτό θα είναι πεντανόστιμο». Το αγόρι χάρηκε που έφτασε τη μπάλα του. Δεν πρόσεξε ότι η γάτα έτρωγε το νόστιμο ψάρι.

Translation in English:
One day there was a playful cat who saw that a butterfly was sitting on a bush. She leaped forward because she wanted to catch it. Meanwhile, a cheerful boy was coming back from fishing holding a bucket and a ball in his hands. He saw that the cat was chasing something. The boy said: "Cat, cat, what are you chasing?". But the butterfly flew away and the cat fell into the bush. The boy was startled and the ball fell out of his hand. But he wanted to get his ball back and decided to catch it. Meanwhile, the cat noticed the boy's bucket and thought: "What might the boy have left in the bucket?". The boy began pulling his ball out of the water. At the same time, the cat was reaching the fish that the boy left and was thinking: "This must be delicious". The boy was happy to have his ball back. He didn't notice that the cat was eating the tasty fish.

<Dog story>

Greek text:
Μια μέρα ήταν ένας παιχνιδιάρης σκύλος που είδε ότι ένα ποντίκι στεκόταν κοντά σε ένα δέντρο. Πήδηξε ψηλά επειδή ήθελε να το πιάσει. Στο μεταξύ ένα χαρωπό αγόρι γυρνούσε από τα ψώνια κρατώντας μια σακούλα και ένα μπαλόνι στα χέρια του. Είδε ότι ο σκύλος κάτι κυνηγούσε. Το αγόρι είπε: «Σκύλε, σκύλε, τι κυνηγάς;» Όμως το ποντίκι έτρεξε μακριά και ο σκύλος έπεσε πάνω στο δέντρο. Το αγόρι ξαφνιάστηκε και του έφυγε το μπαλόνι από το χέρι. Αλλά ήθελε το μπαλόνι του και αποφάσισε να το πιάσει. Στο μεταξύ, ο σκύλος πρόσεξε τη σακούλα του αγοριού και σκέφτηκε: Τι να έχει το αγόρι μέσα στη σακούλα; Το αγόρι άρχισε να τραβάει το μπαλόνι του από το δέντρο. Την ίδια ώρα, ο σκύλος έφτανε το λουκάνικο που άφησε το αγόρι και σκεφτόταν: «Αυτό θα είναι πεντανόστιμο». Το αγόρι χάρηκε που πήρε πίσω το μπαλόνι του. Δεν πρόσεξε ότι ο σκύλος έτρωγε το νόστιμο λουκάνικο.

Translation in English:
One day there was a playful dog who saw that a mouse was standing near a tree. He leaped forward because he wanted to catch it. Meanwhile, a cheerful boy was coming back from shopping holding a bag and a balloon in his hands. He saw that the dog was chasing something. The boy said: "Dog, dog, what are you chasing?". But the mouse ran away and the dog bumped into the tree. The boy was startled and the balloon slipped out of his hand. But he wanted to get his balloon back and decided to catch it. Meanwhile, the dog noticed the boy's bag and thought: "What might the boy have left in the bag?". The boy began pulling his balloon out of the tree. At the same time, the dog was reaching the sausage that the boy left and was thinking: "This must be delicious". The boy was happy to have his balloon back. He didn't notice that the dog was eating the tasty sausage.

Commentary

Time travel in the development of cross-linguistic narrative evaluation

Barbara Zurer Pearson
University of Massachusetts Amherst

The work reported in this book is a welcome and timely contribution for researchers and clinicians everywhere. From the point of view of my personal time machine that travels from where the field of narrative development was when I entered it to here and now, the MAIN is both important and amazing: It is important because of the importance of its topic – stories are central in our lives and crucial for learning. This work is amazing because until recently, it would not have seemed possible to even think about a single instrument that could distinguish typical and atypical narrative development, regardless of the particular language and the number of languages spoken.

Important

The importance of narrative for linguistic and conceptual development has been well established in these past 40 years. Communicating through stories, for all ages, is a key part of the human cultural heritage and for children especially, it is a primary tool for making sense of their world and sharing their experience of it. As Bruner convinces us in *Actual Minds, Possible Worlds* (1986), through stories, children learn how to encode and interpret what is going on around them. Furthermore, the crucial link between oral stories and literacy development that underlies academic success (described in the Introduction) has the rare quality of being reciprocal. The story is both a means and a goal: To improve children's comprehension and production of stories, have them listen to and tell stories. Conversely, with a rich background of oral or written stories, children get progressively better at appreciating what the stories can tell them.

It took the field of child language a few years to turn its attention to narrative development. As is well-known, modern linguistic theory early on, following Chomsky (e.g. 1965), did not concern itself with text. The scientific study of language took

https://doi.org/10.1075/sibil.61.11per

the sentence as the unit of analysis. Children and learnability always figured large in a psycholinguistic theory of grammar. At least one model had to be learnable by children before they could demonstrate the intelligence and coordination to tie their shoes. However, early study of child language focused most attention on how children move from words and parts of words to structured sentences, the topic of Brown's foundational book, *A First Language: The Early Stages* (1973). Still, within just a few years of Brown's book, child language researchers "broke the sentence barrier," seeking tools to explore the structure of connected speech. Labov had given us the minimal definition of a narrative as at least two connected sentences relating events (Labov & Waletzky, 1967; Labov, 1972). The key is "connected." In order to qualify as a narrative, a sequence of two (or more) sentences must have relationships and/ or dependencies that bridge across the sentences. Very soon, Halliday and Hasan (1976) stepped in to provide terms for us to describe the major ways strings of sentences create the relationships necessary to make them stories.

In my time travel, I see Karmiloff-Smith (1979, 1986) foremost among those to explore the boundary between studies of language up to age 5 and those after that "frontier age." Whereas children by age 5 were considered to have significant mastery of the grammar of sentences, Karmiloff-Smith pointed to "organizing spoken text" as the most important aspect of language acquisition after the age of 5. Her own careful experimental work details the subtle syntactic steps in which children move from juxtaposing sentences that cohere using extralinguistic context to a capacity to create cohesive relationships across sentences through language itself. Others in my personal experience of narrative development, like the de Villiers (P. de Villiers, 1991; J. de Villiers, 2001), explored children's developing capacity to represent "other minds" and to reason about their own and others' mental states. They narrowed in on the development of Point of View, taking the perspective of another person (or character), which is the hallmark of mature narratives and which will figure so large in efficient protocols for evaluating emerging narratives.

For me, as I have noted, this is not academic history, it is personal. In retrospection (for this commentary), I realize that my growth as a researcher parallels the growth of the study of narrative. I first heard of and became enthralled with Chomsky's revolution just months after graduating from college in the mid-1960s. A few years later, I followed Brown with only one foot in the field during the years when I stayed home and watched my own children make the journey from babble to fluent speech. By the time I finally completed the Ph.D., Karmiloff-Smith and Bruner had validated the study of narrative and provided me points of entry. The cross-linguistic question would soon be incorporated by Berman and Slobin in a major international project (*Relating Events in Narrative: A Cross-linguistic Developmental Study*, 1994) that spawned many, many follow-on studies of narrative, my own included.

Amazing

So for me, this collection of studies of the MAIN is nothing short of amazing. Because of the strength of the link between oral stories and literacy, the theory of narrative soon became intensely practical. Following Berman and Slobin (1994), the challenge spread around the globe to study developing narratives in many different languages and cultures. Practically speaking, if discourse followed a common set of principles across cultures, then evaluation of discourse might, too. But cross-linguistic "scoring" of narrative was not yet in sight. So that is why the MAIN as demonstrated in this volume is so amazing to me. The attempts at scoring that we tried out in those early days were so primitive and inefficient by comparison, I could never have imagined it.

Berman and Slobin set the course for looking at the same narrative and comparing similar syntactic and semantic bases for its "action-structures" at the same developmental timepoints (ages 3, 5, 9, and adult) in different languages and eventually in different language pairs. Our lab at the University of Miami participated on the periphery of their network. We enthusiastically collected 400 "*Frog Where Are You*" stories, 2 each from 160 bilingual and 80 monolingual second- and fifth-graders and had transcribed and entered them into CHILDES before realizing the size of the challenge to analyze and evaluate them. Without an established metric in two languages or even a "quality continuum" of some sort, we finally had to pose the question for ourselves of what makes a given story good or less good, mature or less mature – and was it different for the Spanish and English stories. Our system (reported in Pearson, 2002) gave us a very detailed picture of the range of strategies pursued within the narrative by the participants in each of their languages, but it was bulky and time-consuming, and worse, it took great efforts to achieve reliability.

My next foray into narrative development was in helping develop the Short Narrative section of the DELV Norm-referenced test (Seymour, Roeper, & de Villiers, 2005). That work was more narrowly directed at assessment and at that point, we took a step in the direction of the MAIN in its generality and efficiency. DELV narrative scoring honed in on minimal criteria for deciding questions of quality – with no need to transcribe, no need for an inventory of elements of a story grammar, or painstaking diagrams of links across sentences. Like the MAIN, it used comprehension as well as production, and focused in its productive items on the child's references to cognitive states and temporal relationships. In comprehension, questions from the examiner probed for whether the child reported the character's activity or the character's thought process, and whether the child's explanation of a pictured event made reference to the character's false belief, or reported the character's goal, or failed to provide a relevant interpretation. With its brevity and simplicity, reliability was a minor issue. The DELV Short Narrative serves well to distinguish typical from atypical development across language varieties (Burns et al., 2012) – but only in one language.

Fast forward a decade and Natalia Gagarina leads the COST Action IS0804 Working Group on Narrative Assessment (Armon-Lotem, de Jong, & Meir, 2015) and effectively ups the ante on what a schema for evaluating narrative development should include. Gagarina and her team, including Bohnacker, her co-author and co-editor for this volume, want – as the whole COST project aimed to do – to create an efficient tool for evaluating stories that can be used for *any* language to compare typical and delayed language development (DLD), AND can be used across languages to compare monolinguals speaking different languages, AND is suited for use in the languages of multilinguals speaking any two or more languages, starting with the 30+ languages represented in the COST project itself. Crucially, the "specifications" for the Working Group's deliverable called for it to be adaptable to any combination of languages, whether closely related or as distant as possible. They chose a slightly different subset of elements than did the DELV narrative test, but essentially, as Bohnacker and Gagarina explain in the Introduction, they focus on a child's appreciation of episodic structure, the distinct points of view of the characters, and how the child integrates them and projects them onto activities and reactions of the characters beyond the story that the child is asked to infer. With its focus on efficiency, the MAIN takes no notice of descriptions of the action sequence or the well-formedness of the child's responses relative to a prescriptive grammar (to the extent that the response is intelligible).

What makes this work even more amazing is its potential for broad application. Bohnacker and Gagarina assure us that the MAIN has already been used with thousands of children, starting with the original COST participants through the 800 more reported in this book – enough for at least two or three standardization studies according to current conventions of test construction. However, those conventions require a reference group that is similar to the group to be evaluated. What could be the representative sample for "every child in the world"? Even every 4- to 6-year-old? Clearly, it does not exist, at least not with our current state of technology (– but I have learned to "never say never"). Given world enough and time, one could imagine MAIN norms being developed for four or five "representative languages," and perhaps some of the major language pairs around the world – Spanish-English, Chinese-English, German-Turkish, French-Arabic, Japanese-Korean, maybe Portuguese and an indigenous language like Wapichana.

Still, the lack of a true norming does not mean that the authors cannot wring more utility from what they have already given us. They can, for example, do what they have done in this volume. Bohnacker and Gagarina follow Karmiloff-Smith in bringing together studies that span the divide at 5 years, and they can add their confirmation (as in Figure 3 of the Introduction) that a jump is indeed observed between ages 4 and 5 years. They have collected studies to illustrate several of the major research and evaluation questions the MAIN can be used to address, especially testing the parameters of the usefulness of its story comprehension protocol. The volume illustrates its use in nine more or less parallel studies with ten languages and even more language pairs. The studies use a variety of research paradigms such that together they provide several models for its use with other, non-tested languages and language pairs.

Best of all, MAIN is in the public domain, so we can anticipate its use everywhere in the world.

References

Armon-Lotem, S., de Jong, J., & Meir, N. (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol: Multilingual Matters. https://doi.org/10.21832/9781783093137

Berman, R., & Slobin, D. (1994). *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674732469

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Burns, F. A., de Villiers, P. A., Pearson, B. Z., & Champion, T. B. (2012). Dialect-neutral indices of narrative cohesion and evaluation. *Language, Speech, and Hearing Services in Schools*, 43(2),132–152. https://doi.org/10.1044/0161-1461(2011/10-0101)

CHILDES, Child Language Data Exchange System. Retrieved from <https://childes.talkbank.org/> (30 June, 2020).

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

de Villiers, P. A. (1991). English literacy development in deaf children: Directions for research and intervention. In J. Miller (Ed.), *Research in child language disorders: A decade of progress* (pp. 349–378). Austin, TX: ProEd.

de Villiers, J. G. (2001). Language acquisition, point of view and possible worlds. Introduction to symposium. In M. Almgren, A. Barrena, M-J. Ezeizabarrena, I. Idiazabal, & B. MacWhinney (Eds.), *Research on child language acquisition: Proceedings of the 8th International Congress of Child Language, San Sebastian*, 1999 (pp. 981–983). Somerville, MA: Cascadilla Press.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. Cambridge: Cambridge University Press.

Karmiloff-Smith, A. (1986). Some fundamental aspects of language development after age 5. In P. Fletcher, & M. Garman (Eds.), *Language acquisition: Studies in first language development* (pp. 455–474). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620683.026

Labov, W. (1972). *Language in the inner city: Studies in the Black English vernacular*. Philadelphia, PA: University of Pennsylvania Press.

Labov, W., & Waletsky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helms (Ed.), *Essays on the verbal and visual arts* (pp. 12–44). Seattle, WA: University of Washington Press.

Pearson, B. Z. (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Clevedon: Multilingual Matters. https://doi.org/10.21832/9781853595721-008

Seymour, H. N., Roeper, T., & de Villiers, J. G. (2005). *DELV-NR (Diagnostic Evaluation of Language Variation) Norm-Referenced Test*. San Antonio, TX: The Psychological Corporation. [Now published by Ventris Learning, Sun Prairie, WI].

Index

Comprehension of texts and understanding of questions is a cornerstone of successful human communication. Whilst reading comprehension has been thoroughly investigated in the last decade, there is surprisingly little research on children's comprehension of picture stories, particularly for bilinguals. This can be partially explained by the lack of cross-culturally robust, cross-linguistic instruments targeting early narration. This book presents an inference-based model of narrative comprehension and a tool that grew out of a large-scale European project on multilingualism. Covering a range of language settings, the book uses the Multilingual Assessment Instrument for Narratives to answer the question which narrative comprehension skills (bilingual) children can be expected to master at a certain age, and explores how such comprehension is affected (or not affected) by linguistic and extra-linguistic factors. Linking theory to method, the book will appeal to researchers in linguistics and psychology and graduate students interested in narrative, multilingualism, and language acquisition.

ISBN 978 90 272 0808 8

9 789027 208088

JOHN BENJAMINS PUBLISHING COMPANY