This is the peer reviewed version of the following article: Valentini, A., & Serratrice, L. (2023). Longitudinal predictors of listening comprehension in bilingual primary school-aged children. Article accepted in Language Learning on 11 March 2022, which has been published in final form at [Link to final article using the DOI].

Longitudinal predictors of listening comprehension in bilingual primary school-aged children

Alessandra Valentini[1,2] & Ludovica Serratrice[1,3]

[1] School of Psychology and Clinical Language Sciences, University of Reading

[2] School of Psychology, University of Surrey

[3] Department of Language and Culture, UiT the Arctic University of Norway

Authors' contribution:

Alessandra Valentini and Ludovica Serratrice were jointly responsible for the conception and the design of the study, for the interpretation of the data, and for drafting and revising the manuscript.

Alessandra Valentini was solely responsible for the data collection and for the data analysis.

Address for correspondence:

Ludovica Serratrice, School of Psychology and Clinical Language Sciences, University of Reading, Harry Pitt Building, Early Gate, Reading RG6 7BE, United Kingdom.

Email: l.serratrice@reading.ac.uk

The authors have no conflict of interest to declare.

Abstract

Research on monolingual children has shown that listening comprehension is predicted by a range of language and cognitive skills; less is known about predictors of listening comprehension in bilingual children and about the role of language input. This study presents longitudinal data on predictors of English listening comprehension in 100 bilingual children between the ages of 5;8 and 6;8. Children were tested three times on their literal and inferential comprehension of stories. Vocabulary, morphosyntax, attention, and memory were included as predictors of listening comprehension alongside a measure of English input. Children showed growth over time in both literal and global inference questions, with performance on local inferences remaining stable over time. Vocabulary depth and morphological knowledge explained listening comprehension abilities in all types of questions, but not their growth. English input had a mediated effect on listening comprehension via morphological knowledge and vocabulary depth, but no direct effect.

Introduction

Listening comprehension, i.e. the ability to understand spoken language, is essential for successful communication and has a prominent role in literacy acquisition and reading comprehension (Hoover & Gough, 1990). Understanding spoken language relies on the ability to retain and store information to create an integrated mental model of the state of affairs, i.e. a situation model (Kintsch, 1994) including information that is mentioned overtly, and information that is only suggested by the text. In making inferences, listeners go beyond what is stated explicitly and make informed guesses about what is implicitly intended. Inferencing skills are thus necessary to make connections between pieces of information in the text (local inferences), or with pre-existing background knowledge outside of the text (global inferences). Local inferences are necessary to integrate two propositions through the mapping of related words, for example between synonyms, category-exemplar pairings, or to resolve anaphoric dependencies. Global inferences, on the other hand, are connections made between information in the text and general background knowledge acquired previously, for example through personal experiences or reading.

A growing body of research has investigated how monolingual children use different language and cognitive skills in listening comprehension (Alonzo, Yeomans-Maldonado, Murphy, Bevens, & LARRC, 2016; Currie & Cain, 2015; de Bree & Zee, 2020; Kim, 2016; Florit, Roch, & Levorato, 2011; 2013; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012; Strasser & Del Rio, 2014). However, we do not yet know how linguistic and cognitive skills predict listening comprehension concurrently and longitudinally in children who speak more than one language. Because of the distributed nature of exposure to their languages, bilingual children also offer a unique opportunity to investigate the role of relative amount of input in the process of listening comprehension.

The present study has three distinct aims. First, we intended to explore the role of foundational cognitive skills (attention and memory), foundational language skills (vocabulary and grammar), and higher order skills (comprehension monitoring and inferencing) in predicting listening comprehension abilities and their growth in bilingual children. Second, we investigated potential differences between literal and inferential comprehension, in terms of their growth and their predictors. Third, we modelled the role of the amount of input in the language of schooling (English) on listening comprehension, considering its possible direct effect, as well as its indirect effect through other language skills.

Predictors of listening comprehension

Predictors of listening comprehension can be broadly grouped into three categories: foundational cognitive skills, foundational language skills, and higher order skills (Kim, 2016).

The foundational cognitive skills most often associated with listening comprehension are memory, attention and IQ (Kim, 2016; Strasser & Del Rio, 2014). Most of the research on memory and listening comprehension has focussed on the effects of short-term memory, i.e. the capacity of the short-term storage, and of working memory (Baddeley, 1986), i.e. the ability to manipulate information from short-term memory (Florit, Roch, Altoé, & Levorato, 2009; LARRC, Jiang, & Farquharson, 2018). Results have been mixed with some studies showing the effects of both measures (Florit et al., 2009; 2013), while others only found an effect of working memory (Silva & Cain, 2015). Given the mixed results, we considered both types of measures in the present research.

Another important cognitive skill related to comprehension is attention, specifically the ability to focus on the task of listening long enough to process the information, and the ability to inhibit irrelevant stimuli. However, while attention skills have been associated with language comprehension (LARRC et al., 2018), some studies have suggested that attention, measured using behavioural checklists, might only have an indirect effect on listening comprehension, via its effect on other language skills (Kim, 2016). In the present study, we focussed specifically on the role of auditory attention, measured directly rather than using a checklist.

When considering foundational language skills, vocabulary and morpho-syntax have repeatedly been associated with language comprehension (Alonzo et al. 2016; Kim, 2016; Strasser & Del Rio, 2014); understanding single words and their structural relationship within a sentence are the essential first steps to comprehending the meaning of a text. When considering vocabulary, recent research has highlighted the importance of distinguishing between breadth and depth of vocabulary knowledge (Ouellette, 2006), and that bilingual children might lag behind monolingual peers in one aspect of vocabulary but not the other (Dixon, Thomson, Fricke, 2020). Vocabulary breadth, defined as the number of entries in the mental lexicon, has been the focus of research in both reading (Eason, Goldberg, Young, Geist, & Cutting, 2012; Muter, Hulme, Snowling, & Stevenson, 2004) and listening comprehension (Silva & Cain, 2015). More recent research has suggested that vocabulary depth, namely the extent of word-related knowledge, and the density of a speaker's semantic network - i.e. the amount of links between words - plays a crucial role in comprehension (de Bree & Zee, 2020; Lepola et al. 2012). In essence, the quality of lexical knowledge operationalized as vocabulary depth predicts comprehension (Perfetti, 2007). Studies that have used both vocabulary breadth and depth tend to report similar (Florit et al., 2009; 2013) or stronger effects of the latter on comprehension (Strasser & Del Rio, 2014). When differentiating vocabulary breadth and depth, it is important to distinguish between tasks that tap into expressive or receptive vocabulary. To avoid the confound of task difficulty, in the present study we used several different measures of vocabulary depth, none of which required the production of definitions, as this can be particularly difficult for bilingual children who often have a gap between receptive and expressive skills (Gibson, Oller, Jarmulowicz, & Ethington, 2012).

Morpho-syntactic knowledge, namely children's ability to comprehend and produce syntactic constructions (e.g. passives vs. actives; subordinates vs. main clauses) and inflectional and derivational morphology (e.g. suffixes for tense changes, or changes to part of speech), has previously been linked to listening comprehension, for example Kim (2016) found a direct effect of syntactic awareness, i.e. the ability to distinguish between grammatical and non-grammatical constructions, on listening comprehension in six- to seven-year-olds, as well as a mediated effect via comprehension monitoring. In Babayiğit (2014) morpho-syntactic skills measured via sentence repetition were significant predictors of listening comprehension for 9 to 10-year-olds. However, not all studies have found an effect of morpho-syntactic knowledge on listening comprehension (e.g. see Alonzo et al., 2016). Overall, while it would be logical to expect a positive relationship between morpho-syntactic knowledge and language comprehension, this relationship may vary as a function of population and assessment method. In this study we included separate measures of syntactic and morphological knowledge to tap into these two relatively separate constructs.

Among higher-order cognitive skills, comprehension monitoring and inferencing are the most widely studied. Comprehension monitoring is the ability to check one's own understanding of a text, and the ability to detect any inconsistencies within the text itself (Ruffman, 1996). Several studies have shown a significant effect of comprehension monitoring on listening comprehension (Kim, 2016) in children as young as 5 (Strasser & Del Rio, 2013).

Making inferences, i.e. the ability to link information within the text (local inferences) or outside of the text (global inferences), is part of the comprehension process itself, however, several studies have considered inference making as a predictor of broader comprehension skills, finding a link between inferencing and listening comprehension (Florit et al., 2011; Kim, 2016). Most studies have employed verbal inferencing tasks where children listen or read passages and answer inferential questions, but a few studies found a way to measure inferencing without relying of children's linguistic abilities, using wordless picture books, finding a correlation between these tasks and reading and listening comprehension (Lepola et al., 2012; Paris & Paris, 2003). Similarly, in the present study we used a wordless picture task to explore the effect of children's inference making abilities independently of their verbal language abilities.

Aspects of listening comprehension

Literal understanding of what is explicitly said in a text, and the ability to make local and global inferences, underpin listening comprehension. However, the predictive power of foundational cognitive and language skills may vary depending on the aspect of comprehension investigated. Literal comprehension requires memory for the details of a story, while local inferences draw on syntactic, semantic and discourse knowledge, and global inferences rely on the comprehender's semantic knowledge, as well as their general world knowledge. Because of these different demands, children tend to find literal questions easier than inferential questions (Eason et al., 2012), and inferential questions that rely on background knowledge are generally harder than those requiring text connections (Barnes, Dennis, & Haefele-Kalvaitis, 1996).

As for predictors of different aspects of comprehension, a few studies have highlighted the importance of vocabulary depth for global inferences rather than for literal comprehension (Cain & Oakhill, 2014; Currie & Cain, 2015). Other studies have shown a reciprocal relationship between vocabulary and the ability to answer inferential questions (LARRC, Currie, & Muijselaar, 2019). Working memory has also been highlighted as a better predictor of inferential than literal questions (Alptekin & Erçetin, 2011). In the present study we differentiate between predictors of literal comprehension, and the ability to make local and global inferences respectively.

Effect of language input on bilingual children's comprehension

Most studies on predictors of listening comprehension have focussed on monolingual participants. When considering bilingual children, several studies have explored predictors of reading comprehension (Bowyer-Crane, Fricke, Schaefer, Lervag, Hulme, 2017; Melby-Lervåg & Lervåg, 2014), while there is little research on predictors of listening comprehension in this population. The few studies that explored this topic showed the importance of vocabulary for listening comprehension (Burgoyne, Kelly, Whiteley, & Spooner, 2009; Hutchinson, Whiteley, Smith, & Connors, 2003). Babayiğit and colleagues (Babayiğit, 2014; Babayiğit & Shapiro, 2020) showed similar levels of predictions of vocabulary (breadth) and grammar measures (sentence repetition and syntactic knowledge) on listening comprehension in their monolingual and bilingual groups aged 9 to 10, and in one study they found an effect of age for the monolingual but not the bilingual group (Babayiğit & Shapiro, 2020). Age could be a proxy for maturation or for amount of language experience, especially for monolingual children, as older children have been exposed to language for longer. The absence of an age effect for bilingual children could suggest that amount of input, rather than age, is a better predictors in bilingual children, as amount of language experience is not just a function of age in this group. Hammer et al. (2012) did show that amount of English exposure significantly predicted Spanish-speaking 5-year-olds' ability to retell a story. However, most studies on listening or reading comprehension in bilingual children do not explicitly model the predictive role of input, even when they document the amount of language exposure in their sample (Babayiğit, 2014; Babayiğit & Shapiro, 2020; Bowyer-Crane et al., 2017).

Language input is one of the strongest predictors of the rate of language development in monolingual and bilingual children (Huttenlocher et al., 1991; Pearson, 2007); measures of input explain variation in vocabulary development in bilinguals (Blom, 2010; Paradis, 2011, Sun et al., 2018) as well as variation in grammar knowledge and its development in both monolinguals (Huttenlocher et al., 2002) and bilinguals (Hoff et al., 2018; Grüter & Paradis, 2014; Thordardottir, 2019). The study of the effects of input in bilingual acquisition is of particular theoretical and practical relevance. Theoretically, individual differences in the amount of language experience in each language makes bilingualism an ideal test case to investigate how much input affects different aspects of language knowledge and their development. Practically, a deeper understanding of the relationship between input, language skills, and their growth, allows practitioners to contextualise expectations for the achievement of bilingual children based on their language background. For these reasons we included a measure of input in our analyses. Our aim was to explore both the direct effect of English input on listening comprehension, as well as its indirect effect via other language skills like vocabulary and grammar.

Research questions

In the light of the literature on listening comprehension in monolingual children, and the relative paucity of research on bilingual children, we investigated the following questions:

1) How do foundational cognitive and language skills predict listening comprehension in bilingual children over time?

On the basis of previous research (Kim, 2016) our hypothesis was that grammar (either syntactic and/or morphological knowledge) and inferential skills would have a direct effect on listening comprehension, over and above other cognitive and language skills. We also expected possible direct effects of vocabulary (either breadth or depth) (de Bree & Zee, 2020; Silva & Cain, 2015) and memory (Kim, 2016).

2) Is the growth of comprehension of literal information and of local and global inferences differentially affected by foundational cognitive and language skills?
   Our hypothesis was that different aspects of listening comprehension might grow differently over time, and that different predictors might explain different comprehension abilities. We specifically expected that vocabulary depth (Cain & Oakhill, 2014; Currie & Cain, 2015) would be a better predictor of local and global inferences than of literal comprehension.
3) How does amount of English input influence listening comprehension in bilingual children?
   Our hypothesis was that English input would affect listening comprehension, but that this effect might be at least partly mediated by a direct relationship with vocabulary (Blom, 2010; Paradis, 2011) and grammar (Hoff et al., 2018; Huttenlocher et al., 2002).

Methods

Data, scripts, and supplementary materials are available at the following URL: https://osf.io/2fa3c/?view_only=cc167f7a3484432b9d5d56e4b87d4032

Participants:

After receiving ethical approval by the research ethics committee at our institution, we recruited 100 bilingual children (48 girls; mean age at fist testing session of 5;8, SD = .29) from 9 schools in the South East of England; 89 participants completed all testing sessions. Data from all children were included in the analyses. Children classified by their school as having English as an Additional Language (EAL) but not otherwise included in the Special Education Needs register were invited to take part in the study. The Department for Education for England defines children with EAL as those children “who were exposed to more than one language (which may include English) during early development”. We decided to include bilingual children according to this broad definition to reflect the linguistic diversity of bi-multilingual children in English primary schools. Children spoke 28 different languages (43% Polish, 7% Hindi, 5% Arabic, 5% French, 5% Romanian, 4% Malayalam, 4% Nepali) and are a representative sample of the current composition of multilingual classrooms in England.

Procedure:

Parents and children gave their consent to participate in the study. Testing sessions were carried out at 3 time points between Year 1 and Year 2 of primary school: Autumn-Spring term (October/February) of Year 1 (Time 1), Spring-Summer term (April/July) of Year 1 (Time 2), and Autumn-Spring term (October/February) of Year 2 (Time 3). Children completed the tasks over two testing sessions at each time point to avoid fatigue. Tasks were completed in the same given order and in the same testing session (first or second session) at all time points by all children.

The data in the present paper are part of a larger longitudinal study in which several language and cognitive measures were collected over time. Children were tested on a series of cognitive and language skills. All tasks were administered in English. Here we report data from the bespoke listening comprehension task completed at all time points, and Time 1 data for the language and cognitive tasks to investigate which language and cognitive measures at the start of the study would predict listening comprehension development over time.

Materials:

Most of the measures used in the present study were tasks from standardised tests used in previous studies with monolingual children. Only raw scores were used in the analyses, as these tests were normed on a monolingual population.

Foundational cognitive skills: Children's general non-verbal abilities were measured using the Matrix Reasoning subtest from the Wechsler Preschool & Primary Scale of Intelligence - Fourth UK Edition (WPSSI – 4 UK – Wechsler, 2013) where children had to choose the best picture to complete a pattern.

Children's attention skills were assessed using two sub-tests of the Test of Everyday Attention for Children (TEA-Ch – Manly, Robertson, Anderson, & Nimmo-Smith, 1998). The Score! subtest asks children to keep count of a series of randomly spaced sounds. The Walk Don't Walk subtest asks them to respond differently to different sounds.

Short-term memory skills were measured using the forward digit span from the Clinical Evaluation of Language Fundamentals - Fourth Edition UK (CELF – 4 UK – Semel, Wiig, & Secord, 2006); working-memory skills were assessed using the backwards digit recall task from the CELF – 4 UK, and a bespoke backwards words recall task where children were asked to repeat a series of words backwards (see Appendix S2). For this task we followed the testing procedure described in Florit et al. (2013).

Foundational language skills: To measure vocabulary breadth we used the British Picture Vocabulary Scale (BPVS –3 - Dunn, Dunn, & NFER, 2009), a word-picture matching task. To assess vocabulary depth we administered three measures. The Synonyms and the

Opposites subtests from the Test of Word Knowledge (TOWK - Wiig & Secord, 1992) requires the child to select the correct synonym or antonym for a given word. Words were presented in writing, and read aloud by the experimenter. Furthermore we administered the Word Classes 1 subtest of the CELF – 4 UK (Semel et al., 2006), where children have to identify conceptually related pictures, and verbalize this relationship. This task yielded a receptive score (number of pairs correctly identified), and an expressive score (number of relationships correctly explained).

For morpho-syntactic knowledge children completed the Word Structure subtest from the CELF – 4 UK (Semel et al., 2006), which requires children to produce the final word in a given sentence with its correct morphological ending, and the short version of the Test of Reception of Grammar (TROG – 2; Bishop, 2003; see Whiteside & Norbury, 2017 for TROG short), a sentence-picture matching task.

Higher-level skills: The non-verbal inferencing task was adapted from story A of the retelling task from the Multilingual Assessment Instrument for Narratives (MAIN – Gagarina et al., 2012). The experimenter pointed at each picture in turn, describing its content, but not giving any information to explain their meaning. After picture presentation, the child was asked the ten comprehension questions of the MAIN. This type of assessment of inferential abilities is similar to the procedure used in previous studies (Lepola et al., 2012; Paris & Paris, 2003).

The comprehension monitoring task followed a testing procedure supported by the literature (Ruffman, 1996), and required children to indicate whether each of 12 short stories "made sense or not". The task was presented as a computer game and children indicated their choice by pressing the relevant button on the keyboard. Six of the stories presented a logical inconsistency, whereby the third sentence contradicted a feature established by the first sentence (see Appendix S3). Children received a point for each story correctly categorised for a maximum score of 12; story order was randomised. This task had a level of reliability (Cronbach's α = .49) in line with previous research (Cain, Oakhill & Bryant, 2004).

Listening comprehension: To assess listening comprehension we used a bespoke measure of listening comprehension; the Understanding of Spoken Paragraphs (USP) subtest from the CELF (Semel et al., 2006) was administered at Time 1 and Time 3 to validate this measure.

The bespoke comprehension task – which consisted of 3 stories at each time point - measured children's comprehension of literal information and their ability to make local and global inferences. Following Freed and Cain (2017) we attempted to alleviate the memory burden for our participants by dividing the stories into three parts with a mean length of 35 words each; children listened to each part of the story through headphones and were asked two or three questions on what they had just heard by the experimenter (see Appendix S5). Answers were given orally and recorded verbatim on the answer sheet. For each story, each child answered two literal and six inferential questions. The inferential questions were divided into three local inferences, where the child was required to make specific anaphoric inferences by connecting different parts of the text, and three global inferences, where they had to integrate the story with their background knowledge. Four independent assessors with PhDs in linguistics were asked to categorise each question as literal, or as requiring a local or a global inference. Any question with less than 100% agreement was further assessed by the authors, and replaced if problematic.

To avoid effects of repeated exposure, 9 different stories were devised for the bespoke listening comprehension measure, three for each time point. Scores were computed as the number of correct answers. Stories were divided in three groups of similar difficulty to be presented at three time points. Classification was initially based on a pilot sample of 13 Year 1 monolingual children. As a second larger sample (40 children in Year 1, of which 26 bilingual) highlighted different difficulty levels between local inference questions included in the stories at Time 2 and at Time 3, outlier questions were eliminated from the analyses, resulting in 7 local inference questions retained at Time 1 and Time 2, and 5 questions at Time 3 (see the Appendix S4 for details of piloting and question selection). The bespoke listening comprehension measure showed good criterion validity (correlation with USP: Time 1: $r = .62$, $p < .001$; Time 3: $r = .75$, $p < .001$) and reliability (Cronbach's α Time 1 = .76; Cronbach's α Time 2 = .82; Cronbach's α Time 3 = .72).

Parental Questionnaire: Parents completed a questionnaire (adapted from Serratrice and De Cat, 2020; Appendix S6) on how much their children heard English and their home language, allowing us to extrapolate a measure of relative amount of English input, and provided demographic information relating to maternal education and SES. As a proxy for SES, we used the highest occupation in the household following the Standard Occupation Classification of the UK Office for National Statistics (https://onsdigital.github.io/dp-classification-tools/standard-occupational-classification/ONS_SOC_occupation_coding_tool.html). This classification yields lower scores for the higher earning occupations, thus scores were reversed in the analyses, for clarity.

To measure English input, we extrapolated a measure that considered the total amount of English input received by the child at Time 1 (questionnaires were completed by the parents between Time 1 and Time 2). Specifically, we computed the total amount of English input as English exposure percentage multiplied by the length of exposure to English. Length of exposure to English was computed as the number of months of exposure to English calculated from age of first exposure (i.e. the difference between age at Time 1 and age of first exposure). To compute English exposure in percentage we asked parents to indicate who spoke to the child, in which language, and how often over the course of the week, using a five-point scale, then converted into percentages (never = 0, rarely = 25%, half of the time = 50%, usually = 75%, always = 100). The amount of input in English was the sum of the number of hours they spent with each interlocutor over the week multiplied by the percentage of time the child heard English from each and then divided by the total number of hours the child was assumed to be awake – assuming 14 hours a day. For the time spent at school (6 hours a day) we assumed that the child only heard English. We recognise this measure might only approximate the total exposure to English for each child, as it assumes no change over the years, but it was considered a good enough approximation, given that collecting measures of relative amount of input at three time points was not possible due to resource limitations and the lack of direct access to the parents.

Data analysis:

We performed four analyses. First, we explored the relationship between measures using correlations, considering percentages of correct responses for listening comprehension.

Second, we used a repeated measures generalized linear model to explore predictors of listening comprehension at all three time points, considering each question separately for each participant, as the dependent measure. This simultaneously allowed us to consider different intercepts for each participant and for each question. All cognitive and language variables that correlated with the dependent variable were considered as predictors in the model, as well as Question Type and Time, and the interactions between Question Type, Time and all other significant predictors.

Third, to explore the relationship between Time and Question Type more in depth, we modelled literal comprehension, local and global inferences separately.

Fourth, we explored the effect of cumulative amount of English input on listening comprehension. Specifically, we performed a mediation analysis to explore whether English input predicted listening comprehension (measured as percentage scores at each time point), either directly, or indirectly via other language skills (acting as mediators). We also considered maternal education, SES and age at Time 1 as potential mediated predictors of listening comprehension, as the literature shows these measures have an effect on language skills, such as vocabulary and morphosyntax (Paradis, 2011; Unsworth, 2016), but none of these measures correlated significantly with listening comprehension at all time points, thus mediation analyses for these measures were not performed. Analyses were performed in R version 4.1.1 (R Core Team, 2021), and we used SPSS for a PCA analysis of vocabulary measures as indicated below.

Results:

See Table 1 for measures' descriptive statistics. To avoid issues of collinearity, measures of memory and vocabulary depth at Time 1 were combined to form 2 factors in a Principle Component Analysis (PCA). For both factors the correlations between the individual measures were significant and higher or equal .30 (see Appendix S1).

No measure showed floor or ceiling effects, except for non-verbal inferencing, which showed a significant positive skew, as a quarter of the participants obtained a score of 9 out of 10. In the bespoke listening comprehension measure children showed progression over time; children also showed higher scores in local inferences (Time 1: *M* = 61%, *SD* = 24, 95% CI = 56.3% – 65.7%; Time 2: *M* = 64%, *SD* = 25, 95% CI = 59% – 69%; Time 3: *M* = 60%, *SD* = 28, 95% CI = 54.2% - 65.8%) than literal questions (Time 1: *M* = 44%, *SD* = 24, 95% CI = 39.3% – 48.7%; Time 2: *M* = 55%, *SD* = 24, 95% CI = 50.1% - 59.9%; Time 3: *M* = 68%, *SD* = 23, 95% CI = 63.2% - 72.8%) or global inferences (Time 1: *M* = 40%, *SD* = 21, 95% CI = 35.9% - 44.1% ; Time 2: *M* = 49%, *SD* = 25, 95% CI = 44% - 54%; Time 3: *M* = 52%, *SD* = 20, 95% CI =47.8% - 56.2%).

Standard scores are not used in the analyses, however, we computed standard scores at Time 1 for a general overview of children's skills. As expected, children's receptive vocabulary skills were lower than monolingual norms, with 9 children with a standard score lower than 70 in vocabulary breadth (BPVS – 3 – Dunn et al., 2009, mean standard score = 87. 17; *SD* = 11.62). However, children performed within the normal range in the test of non-verbal reasoning (Matrix Reasoning from the WPSSI – UK – Wechsler, 2013, mean scaled score = 9.44; *SD* = 2.31). The average Cumulative English Input was 25 months, corresponding to around 2 years equivalent of monolingual exposure; children varied widely in this measure, with 25 children with a Cumulative English Input lower than 12 months, 36 children with a Cumulative English Input between 12 and 30 months (2 years and a half) and 13 with a Cumulative English Input higher than 48 months (4 years).

Table 1 presents correlations between all measures. A restrictive significance level of .01 was applied. As expected, all the measures of foundational language skills were moderately to highly correlated with each other. The memory factor and the measure of non-verbal abilities were also moderately correlated with all the lower-level language measures. The two measures of attention, on the other hand, only showed significant correlation between themselves and memory skills, except a low correlation between Score! and vocabulary depth. Scores on comprehension monitoring were not at chance (W = 2973.50, $p < .001$), but the task showed no significant relationship with other measures, except low correlations with English input, memory and vocabulary breadth. Non-verbal inferencing showed weak, but significant, correlations with all foundational language measures. Maternal education correlated with SES, English input and vocabulary depth. SES correlated with

English input and both measures of vocabulary. English input correlated with all foundational measures of language, while age of the participants did not correlate with any other measure.

Table 1 also shows significant correlations between listening comprehension and non-verbal abilities, memory, vocabulary breadth, vocabulary depth, both measures of morpho-syntactic knowledge and non-verbal inferencing, thus only these measures were further considered in the models. English input was included in mediation models as predictor of both lower level language skills and listening comprehension due to its correlation with these measures. Maternal education was not further included as it only showed a correlation with listening comprehension at Time 3.

Table 1. Descriptive Statistics and spearman rho correlation coefficients between all measures (raw scores) (significance level: $\alpha = .01$ – bold; N = 100 unless specified)

	Mean	95% CI	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 - Maternal Education (N = 89)	3.36	3.19 – 3.53	-	**.68**	.08	**.32**	.06	.08	.08	.07	**.28**	**.37**	.18	**.29**	.11	.04	.16	.10	**.30**
2 – SES (N = 74)	6.43	5.65 – 7.21		-	.19	**.42**	.26	-.06	-.06	.01	**.38**	**.33**	.18	.23	.11	.06	.22	.12	.22
3 - Age Time 1 in months	68.40	67.70 – 69.10			-	.20	.21	.14	.05	.20	.13	.21	.15	.19	.08	.19	.08	.11	.13
4 - English Cumulative input (N = 92)	25.18	21.81 – 28.55				-	.25	.06	.02	**.40**	**.64**	**.47**	**.48**	**.46**	.07	**.27**	**.51**	**.42**	**.37**
5 - Non-Verbal Abilities	14.26	13.62 – 14.90					-	.06	.05	.22	**.32**	**.46**	**.32**	**.42**	.25	.12	**.30**	**.27**	**.32**
6 - Attention: Score!	5.74	5.31 – 6.17						-	**.35**	**.34**	.17	**.26**	.16	.22	.05	.12	.01	.03	.05
7 - Attention: Walk Don't Walk	15.33	14.55 – 16.11							-	**.28**	.15	.23	.19	.05	.10	.07	.09	.12	.01
8 - Memory Factor	0	-0.12 – 0.12								-	**.52**[1]	**.53**[1]	**.55**	**.51**	.19	**.26**	**.50**[1]	**.50**[1]	**.33**[1]
Forward digit span	4.07	3.90 – 4.24																	
Backwards digit span	2.28	2.11 – 2.45																	
Backwards word total	4.00	3.65 – 4.35																	
Backward word span	1.91	1.82 – 2.00																	
9 - Vocabulary Breadth (BPVS)	68.83	65.78 – 71.88									-	**.66**[1]	**.70**	**.69**	**.36**	**.26**	**.69**[1]	**.54**[1]	**.55**[1]
10 - Vocabulary Depth Factor	0	-0.12 – 0.12										-	**.62**	**.66**	**.31**	.25	**.64**[1]	**.57**[1]	**.54**[1]
Synonyms	7.84	6.81 – 8.87																	
Opposites	9.17	8.14 – 10.20																	
Word Classes Receptive	16.07	15.46 – 16.68																	
Word Classes Expressive	10.58	9.74 – 11.42																	
11 – Morpho-syntax: Word Structures	14.53	13.31 – 15.75											-	**.66**	**.41**	.26	**.73**	**.63**	**.64**
12 - Morpho-syntax: TROG Short	22.21	20.64 – 23.78												-	**.35**	.24	**.56**	**.56**	**.62**
13 - Non-verbal inferencing	7.70	7.36 – 8.04													-	.07	**.38**	**.36**	**.44**
14 - Comprehension Monitoring	7.24	6.91 – 7.58														-	.15	.02	.06
15 – List Comp T1 %	48	44.5 – 51.5															-	**.76***[1]	**.68**[1]
16 – List Comp T2 % (N = 94)	56	51.9 – 60.1																-	**.71**[1]
17 – List Comp T3 % (N = 89)	59	55.5 – 62.5																	-

[1]Pearson correlation reported for normally distributed variables

Predictors of listening comprehension at all time points

Mixed-effects models for binomial data were conducted using generalised linear mixed models (Jaeger, 2008), using the function “glmer” from the package “lme4” (Bates et al., 2014), computed with the software R (R Core Team, 2019). Children provided scores for six literal questions at each time point, for seven local inferences at Time 1 and Time 2, for five local inferences at Time 3, and for nine global inference items at each time point. The scoring for each item was binary (1 = correct response; 0 = incorrect response). Our dependent variable was the binary scoring for each item for each participant at each time point. All 100 participants provided scores at Time 1, 94 provided scores at Time 2, and 89 provided scores at Time 3. Time was used as a continuous variable (Grimm, Ram, & Estabrook, 2017), centred around the second time point. As we hypothesized growth in listening comprehension skills over time, an initial model included the random intercepts terms for both participants and question type, to account for different intercepts due to repeated measures on both participants and items, and the random slope of participant over time as well as the fixed effect of time. Comparing this model with a model that postulated no growth (i.e. a model that only included the random intercepts) confirmed linear growth for listening comprehension scores over time (χ^2 (3) = 70.67, p < .001). However, the comparison between a model with and without random slope highlighted no random slope effect (χ^2 (2) = .48, p = .785), suggesting that, while children showed growth in their listening comprehension scores, as shown by the significant fixed effect of time, this growth was similar for all participants. All further analyses retained a random intercept term for both participants and question but no random slope of participant over time (time-only model).

We then compared this time-only model (using pair-wise Likelihood Ratio Test comparisons; Barr, Levy, Scheepers, & Tily, 2013) with a model that additionally included the hypothesised fixed effects: question type (literal *vs.* local inference *vs.* global inference;

literal questions as reference level), non-verbal abilities, memory, vocabulary breath, vocabulary depth, the measures of morpho-syntactic knowledge and non-verbal inferences. All continuous factors were centred around the mean, except the factors created through PCA. This model improved fit compared to the time-only model (χ^2 (9) = 249.17, $p < .001$).

Hypothesised interactions between time and any significant fixed factor were included one at a time in the model with all fixed effects, but none were significant. The interactions between type of question and the measures of individual differences that were significant were separately introduced to test whether different types of questions were differentially predicted by cognitive and language abilities. Nonconvergent models were reduced to significant predictors only, to facilitate convergence. Neither vocabulary depth (χ^2 = 5.21, p = .074), nor word structure (χ^2 = 4.13, p = .127) showed a significant interaction with question type.

To further explore whether the three type of questions developed similarly over time, the interaction between type of question and time was added to the model. The addition of the interaction improved model fit χ^2 (2) =60.25, $p < .001$). The final model is presented in Table 2. Figure 1 represents the interaction between time and question type. The final model showed no issue of collinearity (VIF for Time and the interaction between Time and Question Type < 5, all other VIF < 3).

Table 2. Generalised repeated measures model (given convergence issues the final model did not include Memory, as the effect of this variable was not significant and had the lowest estimated effect size). Marginal R^2_{GLMM} = .17; conditional R^2_{GLMM} = .30.

Factor	Estimate	SE	95 % CI	Estimate odds ratio	z values		χ^2	
					z value	*p*	χ^2	*p*
Intercept	.42	.16	.11 – .72	1.52	2.69	.007*		
Time	.76	.08	.61–.92	2.14	9.64	<.001*	70.72	<.001*
Question type							130.60	<.001*
Literal *vs.* Local	-.08	.02	-.12 – -.04	.92	-3.41	.002*		
Literal *vs.* Global	.15	.03	.09 – .21	1.16	6.03	<.001*		
Local *vs.* Global	.23	.02	.19 – .27	1.26	11.98	<.001*		
Non-verbal abilities	-.03	.06	-.14 – .09	.97	-.46	.642	.28	.597
Vocabulary								
Breadth	.09	.08	-.08 – .26	1.09	1.07	.285	1.15	.283
Depth	.22	.08	.07 – .38	1.25	2.85	.004*	7.41	.006*
Morpho-syntax								
Word structures	.44	.08	.28 – .60	1.55	5.34	<.001*	22.29	<.001*
TROG short	.01	.08	-.14 – .17	1.01	.16	.870	.04	.846
Non-verbal inferences	.06	.06	-.06 – .38	1.06	.98	.329	1.02	.313
Question type*Time							61.22	<.001*
Literal *vs.* Local	.96	13	.70 – 1.22	2.62	7.70	<.001*		
Literal *vs.* Global	.37	.10	.17 – .57	1.44	3.53	.012*		
Local *vs.* Global	-.60	.12	-.84 – -.36	.55	-4.98	<.001*		

Random effects	Var	SD
Subject	.19	.43
Question	.40	.63

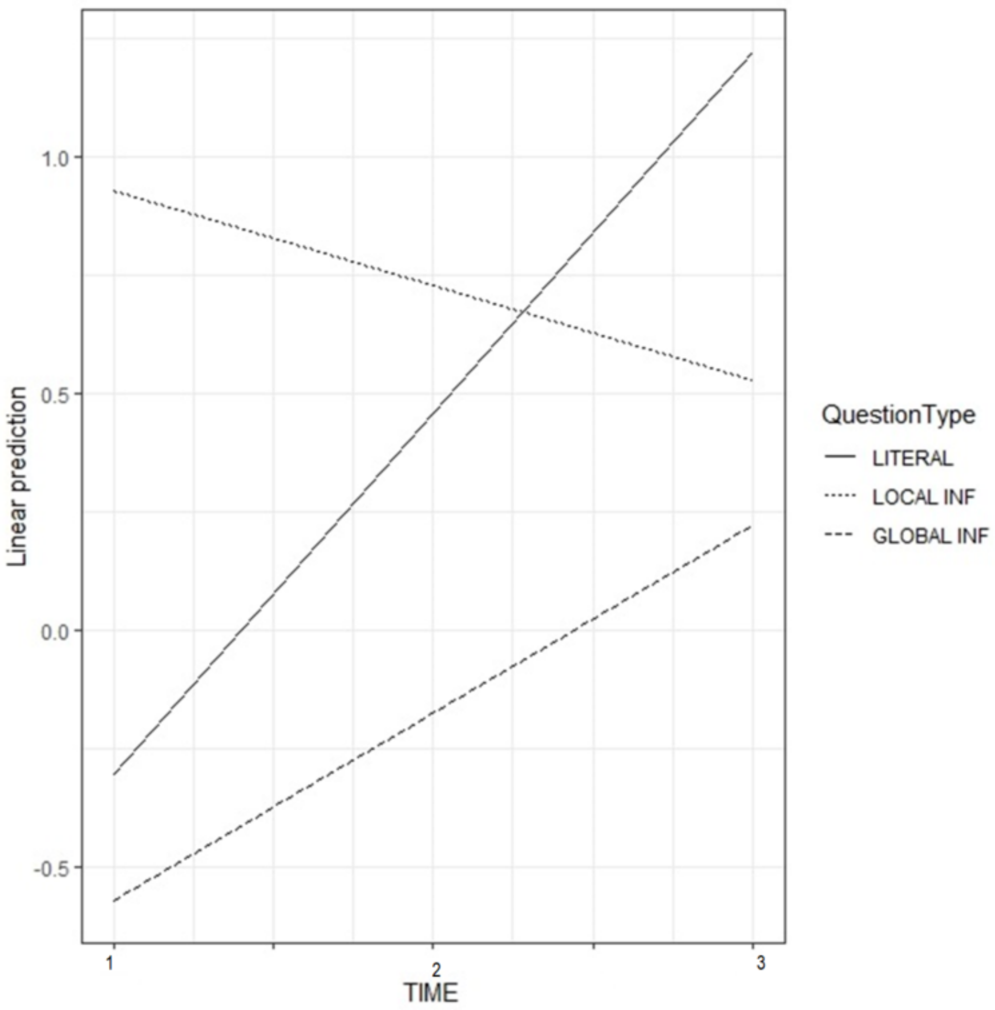

Figure 1. Linear prediction of listening comprehension by time and question type

As shown in Table 2, vocabulary depth and morphological knowledge measured through the Word Structures test significantly predicted listening comprehension. Question type was also a significant predictor, with local inferences emerging as the easiest questions, above literal questions and global inference questions which emerged as the most difficult type of questions. The results also highlight different trajectories for the different questions (Figure 1) with higher growth over time for literal questions, followed by growth in global inferences questions, and a levelling off over time for local inferences questions. The model explained 30% of variance in listening comprehension (this level of explained variance is to be expected for logistic models); of this, 17% was explained by the fixed effects. While low, this value is much higher than the variance explained by Time in the time-only model (1%).

Predictors and differences in growth for each question type

Given the significant interaction between time and question type, three models, one for each type of questions were computed. These models were used to explore growth for each type of question. The same predictors as the ones used in the main model were entered to ensure comparability with the main model. These analyses were deemed appropriate given the significant interaction between question type and time.

The models (see Table 3) show that performance for literal questions and global inference questions, but not for local inference questions, significantly improved over time. The performance for both types of inference questions was predicted by both vocabulary depth and morphological knowledge as in the main model, while only morphological knowledge predicted literal questions. Figure 2 shows participants performance on literal questions, local and global inference questions when children were grouped by morphological knowledge or vocabulary depth. Higher ability children show higher scores but relatively similar growth (or lack of growth) over time compared to low ability children in all question types, except for an apparent lack of growth in global inference questions between Time 2 and Time 3 for the top quartile group in morphological knowledge and the two higher groups for vocabulary depth.

Table 3. Repeated measures mixed effect models for the three types of questions.

	Literal Questions					Local Inferences					Global Inferences				
				z values					z values					z values	
Factor	**Est.**	**95 % CI**	**Est. odds ratio**	**z**	***p***	**Est.**	**95 % CI**	**Est. odds ratio**	**z**	***p***	**Est.**	**95 % CI**	**Est. odds ratio**	**z**	***p***
Intercept	.67	-.13 – 1.47	1.95	1.64	.101	.58	.28 – .88	1.79	3.84	<.001*	-.13	-.55 – .29	.88	-.61	.539
Time	1.12	.90 – 1.34	3.07	10.04	<.001*	-.03	-29 – .24	.97	-.20	.840	.40	.25 – .56	1.49	5.05	< .001*
Non-verbal abilities	-.04	-.22 – .14	.96	-.45	.652	-.01	-.15 – 15	1.00	-.02	.982	-.04	-.18 – .11	.96	-.51	.613
Memory	.04	-.17 – .24	1.04	.35	.727	.01	-.16 – .19	1.01	.16	.872	-.03	-.19 – .14	.97	-.33	.744
Vocabulary															
Breadth	-.11	-.36 – .14	.89	-.89	.375	.13	-.09 – .35	1.14	1.13	.260	.20	-.01 – .40	1.22	1.90	.057
Depth	.14	-.10 – .37	1.15	1.14	.256	.27	.07 – .48	1.31	2.58	.009*	.27	.08 – .46	1.31	2.73	.006*
Morpho-syntax															
Word structures	.67	.42 – .93	1.96	5.07	<.001*	.51	.28 – .73	1.66	4.44	<.001*	.32	.11 – .53	1.37	3.01	.003*
TROG short	.10	-.14 – .34	1.11	.82	.415	-.06	-.26 – .15	.95	-.53	.599	.01	-.19 – .21	1.01	.10	.921
Non-verbal inferences	.03	-.14 – .20	1.03	.35	.730	.08	-.07 – .23	1.08	1.04	.297	.07	-.07 – .21	1.07	1.00	.319

Random effects	**Var**	**SD**	**Marg R^2_{GLMM}**	**Cond R^2_{GLMM}**	**Var**	**SD**	**Marg R^2_{GLMM}**	**Cond R^2_{GLMM}**	**Var**	**SD**	**Marg R^2_{GLMM}**	**Cond R^2_{GLMM}**
Subject	.18	.42	.20	.52	.18	.42	.15	.25	.19	.43	.12	.32
Question	.40	.63			.28	.53			.40	.63		

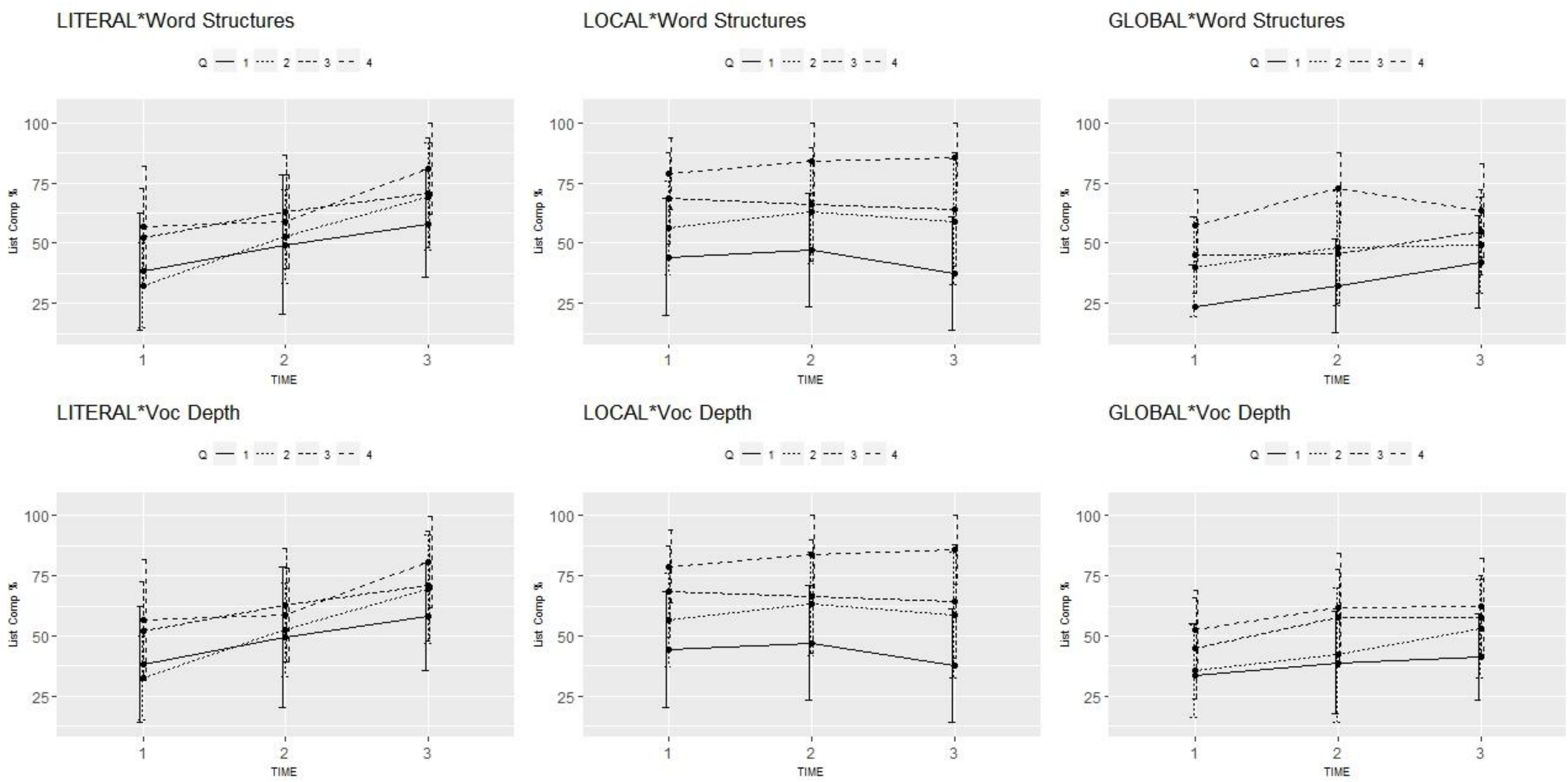

Figure 2. Listening comprehension % scores distribution (with SD) for each question type (literal, local and global inferences from left to right) at each time point (within graphs). Top row: participants divided by score in Word Structures in quartiles (Q). Bottom row: participants divided by score in Vocabulary Depth (Voc Depth) in quartiles (Q).

Mediation Model

English input was the only environmental measure (among English input, maternal education, and SES) to significantly correlate with listening comprehension at all time points. English input also correlated with all other language measures (Table 1). To explore the effect of English input on listening comprehension, by accounting for its possible effect on other language measures, a mediation model was computed. We specifically considered as potential mediators those language factors that were significant in the main mixed-effects model (Word Structures and Vocabulary Depth). The mediation model was built using the R package lavaan (Rosseel, 2012). As dependent variable in the model we used a latent variable directly estimated within the model, computed considering percentages of correct responses in listening comprehension at each time point. Regression paths between listening comprehension at Time 1 and Time 2, and between Time 2 and Time 3 were estimated to further consider the growth of listening comprehension between time points shown in the main model. Fit for the model was computed using maximum likelihood estimation with robust standard errors (MLMV).

The model (Figure 3) explained 78.2% of the variance in listening comprehension. It highlighted no significant direct effect of English input on listening comprehension, but significant indirect effects of English input through Word Structures, and vocabulary depth, as well as significant effects of the mediators, confirming the results of the model in Table 2 (Word Structures: $\beta = .63$, $p < .001$;Vocabulary Depth: $\beta = .29$, $p = .005$). The total effect of English input on listening comprehension was c_1: $\beta = .50$, $p < .001$.

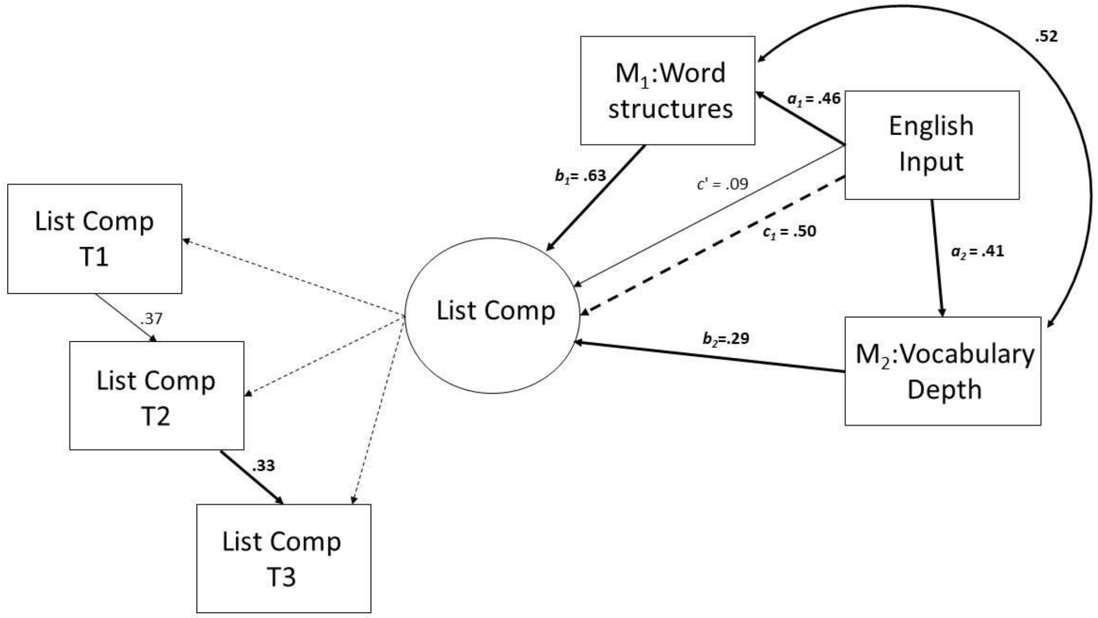

Figure 3. Multiple mediation model with two mediators: M_1, Word Structures and M_2, Vocabulary Depth (χ^2 (4) = .93, *p* = .920, CFI > .999, TLI > .999, RMSEA < .001, SRMR = .011). English input has no significant direct effect (path c': β = .09, 95% CI = -.07 – .23, *p* = .263). English input has a significant indirect effect on Listening Comprehension through M_1, computed as the product of the two paths linking English input on Listening Comprehension through that mediator, that is, a_1b_1 (β = .29, 95% CI: .15 – .44, *p* = .001); similarly, English input has a significant indirect effect on Listening Comprehension through M_2 (Vocabulary Depth), defined as a_2b_2 (β = .12, 95% CI: .03 – .21, *p* = .017). The total effect of English input on Listening Comprehension was computed as: $c_1 = c' + a_1b_1 + a_2b_2$ (β = .50, 95% CI: .33 – .66, *p* < .001).

Discussion

The aim of the current study was to investigate how foundational cognitive and language skills predict literal and inferential listening comprehension over time in bilingual children between the ages of 5 and 7 (Year 1 to Year 2 of primary school in the English school system), and how this relationship is mediated by English input. We also addressed whether the comprehension of literal information and of local and global inferences is differentially affected by foundational cognitive and language skills in bilingual children.

Predictors of listening comprehension

Our results clearly show the importance of vocabulary depth and morphological knowledge in listening comprehension: of all the predictors only these two significantly contributed to explaining the variance in listening comprehension abilities. The importance of measures of vocabulary and grammar is in line with previous research (Kim, 2016; Silva & Cain, 2015). These two variables had a direct effect on listening comprehension, but their interaction with time was not significant, thus vocabulary depth and morphological knowledge predicted listening comprehension abilities overall (i.e. the intercept of a growth model) but not change over time. Vocabulary depth and morphological skills also predicted all three types of comprehension questions similarly (literal, local and global), although vocabulary depth failed to reach significance in the model for literal comprehension. Differently from some of the previous research (Kim, 2016), our study did not highlight a direct effect of memory or non-verbal inferencing skills on listening comprehension, once all other predictors were considered.

In terms of vocabulary knowledge, our findings suggest that the quality of lexical representations (vocabulary depth) is more informative for listening comprehension than the sheer number of words children know (vocabulary breadth). This result is in line with the result of Strasser and Del Rio (2014), who showed that the effect of vocabulary breadth tends to be fully mediated by vocabulary depth in the concurrent prediction of listening comprehension of pre-school children. The importance of vocabulary depth over vocabulary breadth is also in line with the Lexical Quality Hypothesis (Perfetti & Hart, 2002), as vocabulary depth tasks tend to tap into lexical quality, whereas vocabulary breadth task do not. In Perfetti and Hart's view, the link between different aspects of word knowledge is extremely important. They defined a representation of high quality as one that includes information regarding different features of the same word, i.e. the representation of form and meaning, as well as morphological information and information regarding its use, and they suggested that a high quality representation of a word allows children to rapidly access all features of a given word. High quality lexical representations facilitate word processing thus freeing processing resources, and they also provide detailed and extensive semantic information that is necessary for local and global inferences.

The other significant predictor of listening comprehension in all of our analyses is morphological knowledge. The measure of morphological knowledge adopted in this study is heavily linked to word knowledge, and it captures children's ability to manipulate inflectional and derivational endings and to use appropriate anaphoric forms. The importance of morphological knowledge for local inferences, especially those that require linking pronouns with their antecedents, is relatively straightforward. Children rely on their knowledge of pronouns to solve anaphoric local inferences; however the results suggest a more widespread effect of morphological knowledge on all types of listening comprehension questions. The importance of morphological knowledge is in line with previous research (Babayiğit, 2014), although in many previous studies the effect of morphology is often conflated with that of syntactic knowledge (e.g. Babayiğit, 2014; Muter et al., 2004) and has not been consistently replicated (Alonzo et al., 2016). Well-developed morphological knowledge contributes to higher lexical quality (Perfetti & Hart, 2002), and it is a powerful tool that children can rely on to make those text connections leading to successful inferences. In the present study, morphological knowledge emerged as a better predictor than syntactic knowledge. As in previous research (Florit et al. 2013), we found that a sentence-picture matching task was not associated with text comprehension.

Our results also show that, while vocabulary depth and morphological knowledge both have an effect on listening comprehension, they do not seem to affect its growth over time. This result is similar to that of Lepola et al. (2012), who found an effect of vocabulary on listening comprehension concurrently in monolingual Finnish 4-year-olds, but not longitudinally two years later. In their path analysis, which included the autoregressor of listening comprehension, vocabulary explained the level of listening comprehension at Time 1, but did not further explain its growth from Time 1 to Time 3 above what was originally explained by the autoregressor. Similarly, Proctor and colleagues (2012) showed an effect of both vocabulary breadth and morphological knowledge in predicting initial status, but not change in reading comprehension of monolingual and bilingual children in primary school over six months. While vocabulary and morphological knowledge are important in determining children's performance in listening comprehension tasks, they do not necessarily predict developmental progress. As we did not consider growth in vocabulary and morphological skills over time in the present paper, we cannot conclude whether growth in either of these skills might predict growth in listening comprehension. While vocabulary depth was significant in the main model, further analyses failed to show an effect of vocabulary depth on literal comprehension. A reduced role of vocabulary on literal comprehension compared to inferencing is in line with previous research (Cain & Oakhill, 2014). A large and deep lexicon may be more important to make connections within the text rather than for verbatim recall as is the case in literal comprehension. Breadth and depth of vocabulary can also assist in the making of global inferences that require extensive background knowledge. For example, given a scene where people swim in the water and build sandcastles, children who have a deeper and more connected semantic network will have a link between the words sandcastle, water and beach, which will facilitate the inference that the scene takes place on a beach.

A striking result of the present research was the lack of individual growth slopes in listening comprehension over time. Research on growth in comprehension skills doesn't always include random slopes (Raudszus, Segers, & Verhoeven, 2021), as the inclusion of individual difference measures and the consideration of the interaction between time and these measures might account for any individual variance in growth. Our result could be interpreted by considering that we tested children who had started attending primary school. Starting school will reduce the individual variation in the amount of input bilingual children receive in each language, as they will all be exposed to the same amount of English during school hours. English input at school would therefore act as a leveller for these bilingual children whose exposure to the language prior to the start of formal education varied widely. Additional support for this hypothesis comes from the absence of individual differences in the growth patterns in vocabulary breadth in the same sample of children (Valentini & Serratrice, 2021).

Another important result of the present study is the lack of a significant association between all other skills measured and listening comprehension, once vocabulary depth and morphological knowledge were accounted for. Two of these measures in particular, memory skills and inferential skills have been previously shown to be directly associated with listening comprehension (Kim, 2016). The direct effect of memory has not been replicated in all studies, and could be explained by considering that memory might have an indirect association with listening comprehension skills, through its association with other foundational skills (Cain et al., 2004; de Bree & Zee, 2020). The difference between our study and previous ones (Florit et al., 2011; Kim, 2016) regarding the lack of influence of inferencing skills on listening comprehension might be ascribed to the specific measures used. Previous research that found a more prominent effect of inferencing skills on listening comprehension used inferencing tasks that heavily relied on children's language abilities. The similarities between the two tasks might therefore have increased the likelihood of finding a direct relationship between inferencing and listening comprehension. In the present study, however, we specifically chose a measure of non-verbal inferencing that allowed us to measure our bilingual children's inferencing skills without an excessive reliance on their oral language abilities. This allowed us to explore the relationship between inferencing skills and listening comprehension more directly. It is also possible that, by measuring foundational language skills more thoroughly (vocabulary breadth and depth, and morphological and syntactic abilities), part of the variability explained by inferencing skills tasks in other studies was accounted for by other language tasks in the present one. This would further confirm the importance of including measures of vocabulary depth and morphological knowledge when exploring predictors of listening comprehension.

Differences between types of comprehension questions

Literal comprehension and the ability to answer global inference questions grew over time, while the ability to answer local inference questions remained relatively stable. Furthermore, local inferences emerged as the easiest type of questions, followed by literal questions and by global inference questions. This result seems at odds with previous literature reporting literal questions as the easiest types of questions (Alptekin & Erçetin, 2011; Cain & Oakhill, 2014). This difference might be due to the type of questions used, and the level of detail required in the present study: all the information necessary to answer literal question was directly provided in the text, but this information needed to be correctly encoded and retrieved to answer these questions. It is possible that our test of comprehension required a higher level of detail than in other studies, thus making literal questions harder. On the other hand, the difference between local and global inferences, and the emergence of global inference questions as the more complex types of comprehension questions is in line with previous research with monolingual children when reading (Cain & Oakhill, 2014), but not when listening (Currie & Cain, 2015; Freed & Cain, 2017). Global inference questions require the ability to use previous knowledge to extract a deeper meaning from a given text than a literal interpretation might provide. This process is demanding, and this task might be particularly hard for bilingual children in their second language (Hara & Tappe, 2016; Schönpflug & Küpping-Faturikova, 2020). They need to retrieve information that might not have been stored in their second language, or that might have been encoded differently in the two languages (Adams, 2016).

As for local inferences, while children's performance was not at ceiling at Time 1, it is possible that the results at the first time point were already so high that children were left with relatively little room for improvement. It is also possible that the skills required to answer local inference questions develop before the time window examined, and that no major changes take place between 5 and 7 years of age. Almost half of the questions categorised as local inferences (7 out of 19) required the anaphoric resolution of a third person singular subject pronoun (*he, she*). English-speaking monolingual children as young as 3 can use gender information to find the correct antecedent of a subject pronoun (Arnold, Brown-Schmidt, & Trueswell, 2007), but little is known about bilingual children's anaphora resolution in English (see Serratrice & Hervé, 2015 for an overview). Success on local inferences in the present study required children to find an antecedent for an anaphora, either a noun or a pronoun, in the preceding sentence. Whenever the children had to find an antecedent for a pronoun, knowledge of grammatical gender would unambiguously lead them to the correct choice. Similarly, when finding an antecedent for a nominal anaphora there was only one semantically plausible antecedent. Anaphora resolution can be a very complex task when contextual and semantic information increase the ambiguity of a potential antecedent, but it can also be a rather mechanistic process when there is little ambiguity, as in the stories used in the present study. Conversely, the ability to make global inferences requires children to recruit information from long-term memory that grows as a function of their increasing experience of the world, and therefore growth over time is to be expected.

The effect of English input

A key result of the present paper is the effect of the cumulative amount of English input on listening comprehension. Specifically, English input showed a significant, but indirect effect on listening comprehension mediated by morphological knowledge and vocabulary depth. This result confirms the importance of language input for foundational language skills, such as vocabulary and morpho-syntactic abilities (Hoff et al., 2018; Brinchmann, Braeken, & Lyster, 2019). Cumulative amount of English input for the bilingual children in this study did not directly affect listening comprehension, but it directly predicted levels of vocabulary and grammar skills, that, in turn, positively influenced children's ability to comprehend texts.

Limitations

A limitation of the present research is that, contrary to previous models (Kim, 2016) we did not consider possible mediation effect of higher language abilities on the relationship between listening comprehension and cognitive skills. Our results show the importance of vocabulary depth and morphological knowledge in predicting listening comprehension, over and above the influence of any other predictor. However, it is possible that lower level cognitive skills, such as memory and attention, might have a more subtle relationship with listening comprehension, mediated via a relationship between these skills and vocabulary and grammar skills. It could be, for instance, that children with better working memory might be better word learners, with better vocabulary skills, and better vocabulary skills will positively affect their listening comprehension. Our initial choice of a simpler model that did not consider this direction of influence was motivated by the relatively lower number of participants in our study compared to studies that considered these relationships (de Bree & Zee, 2020; Kim, 2016). We believe our model is of value in highlighting the specific importance of vocabulary and morphological skills in predicting listening comprehension, however we cannot rule out the possible (mediated) effects of other cognitive skills.

Another limitation of the present research is the lack of longitudinal measurement of the predictors of listening comprehension as well as English input; specifically, the present model considers how children's abilities at the beginning of the study relate to their listening comprehension longitudinally. This approach was chosen due to lack of longitudinal data on some of the predictors. It is possible that, while initial skills might only predict level of listening comprehension, considering growth in these skills over time might have also predicted growth in listening comprehension. Our study is still novel in attempting to measure many of the possible predictors of growth in listening comprehension, as many of the previous studies did not consider listening comprehension longitudinally (de Bree & Zee, 2020; Kim, 2016), or restricted their analysis to some predictors (Lepola et al., 2012) or both (Florit et al., 2011). However, we acknowledge that including longitudinal measures of the predictors, as well as longitudinal changes in English input might have accounted for more variability in listening comprehension, especially in relation to its growth over time.

Conclusion

Vocabulary depth and morphological knowledge were the most significant predictors of English listening comprehension in bilingual children in the first two years of formal schooling in our study. These skills specifically determined children's performance, but not their growth in listening comprehension abilities over time. The amount of English input children received had a significant impact on their listening comprehension performance, but its predictive power was mediated by its effect on foundational language skills, particularly morphological knowledge and vocabulary depth. These results make a novel contribution to a better understanding of the determinants of listening comprehension in bilingual children and they have pedagogical implications. Increasing children's high-quality lexical representations, specifically providing them with more information about words' meaning and their use, as well as increasing their knowledge of morphology, is likely to have a positive cascading effect on spoken language comprehension.

Regarding different kinds of comprehension questions, the results confirm that global inferences are the hardest kinds of questions for school-aged children, and that both their ability to answer global inference questions and their literal comprehension grow over time.

(9802 words)

References:

Adams, A. M. (2016). How language is embodied in bilinguals and children with specific language impairment. *Frontiers in Psychology*, *7*, 1209. https://doi.org/10.3389/fpsyg.2016.01209

Alonzo, C. N., Yeomans-Maldonado, G., Murphy, K. A., Bevens, B., & LARRC (2016). Predicting second grade listening comprehension using prekindergarten measures. *Topics in Language Disorders*, *36*(4), 312-333. https://doi.org/10.1097/tld.0000000000000102

Alptekin, C., & Erçetin, G. (2011). Effects of working memory capacity and content familiarity on literal and inferential comprehension in L2 reading. *Tesol Quarterly*, *45*(2), 235-266. https://doi.org/10.5054/tq.2011.247705

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22*(4), 527-565. https://doi.org/10.1080/01690960600845950

Babayiğit, S. (2014). The role of oral language skills in reading and listening comprehension of text: A comparison of monolingual (L1) and bilingual (L2) speakers of English language. *Journal of Research in Reading*, *37*(S1), S22-S47. https://doi.org/10.1111/j.1467-9817.2012.01538.x

Babayiğit, S., & Shapiro, L. (2020). Component skills that underpin listening comprehension and reading comprehension in learners with English as first and additional language. *Journal of Research in Reading*, *43*(1), 78-97. https://doi.org/10.1111/1467-9817.12291

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Barnes, M. A., Dennis, M., & Haefele-Kalvaitis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology*, *61*(3), 216-241. https://doi.org/10.1006/jecp.1996.0015

Bates D., Maechler M., Bolker B., &Walker S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, http://CRAN.R-project.org/package=lme4

Bishop, D. V. M. (2003). *Test for the Reception of Grammar – Version 2*. London, UK: Harcourt Assessment.

Blom, E. (2010). Effects of input on the early grammatical development of bilingual children. *International Journal of Bilingualism*, *14*(4), 422-446. https://doi.org/10.1177/1367006910370917

Bowyer-Crane, C., Fricke, S., Schaefer, B., Lervåg, A., & Hulme, C. (2017). Early literacy and comprehension skills in children learning English as an additional language and monolingual children with language weaknesses. *Reading and Writing*, *30*(4), 771-790. https://doi.org/10.1007/s11145-016-9699-8

Brinchmann, E. I., Braeken, J., & Lyster, S. A. H. (2019). Is there a direct relation between the development of vocabulary and grammar?. *Developmental Science*, *22*(1), e12709 https://doi.org/10.1111/desc.12709

Burgoyne, K., Kelly, J. M., Whiteley, H. E., & Spooner, A. (2009). The comprehension skills of children learning English as an additional language. *British Journal of Educational Psychology*, *79*(4), 735-747. https://doi.org/10.1348/000709909x422530

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension?. *L' Annee Psychologique*, *114*(4), 647-662. https://doi.org/10.4074/s0003503314004035

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. https://doi.org/10.1037/0022-0663.96.1.31

Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology*, *137*, 57-75. https://doi.org/10.1016/j.jecp.2015.03.005

de Bree, E., & Zee, M. (2020). The unique role of verbal memory, vocabulary, concentration and self-efficacy in children's listening comprehension in upper elementary grades. *First Language* 1 - 25. https://doi.org/10.1177/0142723720941680

Dixon, C., Thomson, J., & Fricke, S. (2020). Language and reading development in children learning English as an additional language in primary school in England. *Journal of Research in Reading*, *43*(3), 309-328. https://doi.org/10.1111/1467-9817.12305

Dunn, L. M., Dunn, D. M., & NFER. (2009). *British Picture Vocabulary Scale - 3rd ed. (BPVS –3)*. London: GL Assessment Ltd.

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader–text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology*, *104*(3), 515. https://doi.org/10.1037/a0027182

Ehrlich, M.-F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing*, *11*, 29–63. https://doi.org/10.4324/9780203053324-15

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

Freed, J., & Cain, K. (2017). Assessing school-aged children's inference-making: the effect of story test format in listening comprehension. *International Journal of Language & Communication Disorders*, *52*(1), 95-105. https://doi.org/10.1111/1460-6984.12260

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Valimaa, T., Balciuniene, I., Bohnacker, U., & Walters, J. (2012). ˙ MAIN: Multilingual Assessment Instrument for Narratives. *ZAS papers in linguistics* 56. Berlin: Zentrum fur Allgemeine Sprachwissenschaft.

Gibson, T. A., Oller, D. K., Jarmulowicz, L., & Ethington, C. A. (2012). The receptive-expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. *Bilingualism (Cambridge, England)*, *15*(1), 102. https://doi.org/10.1017/s1366728910000490

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: Structural equation and multilevel approaches*. London: The Guiford Press.

Grüter, T., & Paradis, J. (Eds.). (2014). *Input and experience in bilingual development* (Vol. 13). John Benjamins Publishing Company.

Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., & Goldstein, B. (2012). Predicting Spanish–English bilingual children's language abilities. *Journal of Speech, Language, and Hearing Research*, *55*, 1251–1264. https://doi.org/10.1044/1092-4388(2012/11-0016)

Hara, A., & Tappe, H. (2016). Inference generation and text comprehension in bilingual children: A case study. *Literator*, *37*(2), 1-12. https://doi.org/10.4102/lit.v37i2.1287

Hoff, E., Quinn, J. M., & Giguere, D. (2018). What explains the correlation between growth in vocabulary and grammar? New evidence from latent change score analyses of simultaneous bilingual development. *Developmental Science*, *21*(2), e12536. https://doi.org/10.1111/desc.12536

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127-160. https://doi.org/10.1007/bf00401799

Hutchinson, J. M., Whiteley, H. E., Smith, C. D., & Connors, L. (2003). The developmental progression of comprehension-related skills in children learning EAL. *Journal of Research in Reading*, *26*(1), 19-32. https://doi.org/10.1111/1467-9817.261003

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology*, *27*(2), 236. https://doi.org/10.1037/0012-1649.27.2.236

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, 45(3), 337-374. http://dx.doi.org/10.1016/S0010-0285(02)00500-5

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446. https://doi.org/10.1016/j.jml.2007.11.007

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). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294. https://doi.org/10.1037/0003-066x.49.4.294

LARRC, Currie, N. K., & Muijselaar, M. M. (2019). Inference making in young children: The concurrent and longitudinal contributions of verbal working memory and vocabulary. *Journal of Educational Psychology, 111(8)*, 1416-1431. https://doi.org/10.1037/edu0000342

LARRC, Jiang, H., & Farquharson, K. (2018). Are working memory and behavioral attention equally important for both reading and listening comprehension? A developmental comparison. *Reading and Writing*, *31*(7), 1449-1477. https://doi.org/10.1007/s11145-018-9840-y

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., & De Ley, L. (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language, 42*(4), 545–570. https://doi.org/10.1006/jmla.1999.2695

Manly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith I. (1998). *Test of Everyday Attention for Children (TEA-Ch)*. London: Pearson.

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological Bulletin*, *140*(2), 409. https://doi.org/10.1037/a0033890

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology*, *40*(5), 665-681. https://doi.org/10.1037/0012-1649.40.5.665

Office for National Statistics (n.d.) ONS Occupation Coding Tool. Retrieved March, 2019, from https://onsdigital.github.io/dp-classification-tools/standard-occupational-classification/ONS_SOC_occupation_coding_tool.html

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554. https://doi.org/10.1037/0022-0663.98.3.554

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

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. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, *28*(3), 399-410. https://doi.org/10.1017/s014271640707021x

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357-383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of Functional Literacy* (pp. 189–213). Amsterdam: John Benjamins.

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Reading and Writing*, *25*(7), 1635-1664. https://doi.org/10.1007/s11145-011-9336-5

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/

Raudszus, H., Segers, E., & Verhoeven, L. (2021). Patterns and predictors of reading comprehension growth in first and second language readers. *Journal of Research in Reading*, 44(2), 400-417. . https://doi.org/10.1111/1467-9817.12347

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Ruffman, T. (1996). Reassessing children's comprehension-monitoring skills. In Cornoldi, C. and Oakhill, J. (Eds.) *Reading comprehension difficulties: Processes and intervention* (pp. 33-67). New Jersey: Lawrence Erlbaum.

Schönpflug, U., & Küpping-Faturikova, L. (2020). Bilingual children's story comprehension across language boundaries: does retelling improve comprehension?. *International Journal of Bilingual Education and Bilingualism*, *23*(6), 657-675. https://doi.org/10.1080/13670050.2018.1479727

Semel, E., Wiig, E., & Secord, W. (2006). *Clinical Evaluation of Language Fundamentals–4th ed. (CELF - 4 UK)*. London: Pearson.

Serratrice, L. & Hervé, C. (2015). Referential expressions in bilingual acquisition. In L. Serratrice, & S. Allen (Eds), *The Acquisition of Reference* (pp. 311-333). Amsterdam: John Benjamins.

Serratrice, L., & De Cat, C. (2020). Individual differences in the production of referential expressions: The effect of language proficiency, language exposure and executive function in bilingual and monolingual children. *Bilingualism: Language and Cognition*, *23*(2), 371-386. https://doi.org/10.1017/s1366728918000962

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

Strasser, K., & Río, F. D. (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

Sun H., Yin B., Amsah F., O'Brien B. A. (2018b). Differential effects of internal and external factors in early bilingual vocabulary learning: the case of Singapore. *Applied Psycholinguistics*, *39*(2), 383–411. https://doi.org/10.1017/S014271641700039X

Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, 23(1), 236-255. https://doi.org/10.1177/1367006917722418

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Language and the human lifespan series. Bilingualism across the lifespan: Factors moderating language proficiency* (p. 103–121). American Psychological Association.

Valentini, A. & Serratrice, L. (2021). What can bilingual children tell us about the developmental relationship between vocabulary and grammar? *Cognitive Science*, 45(11), e13062. https://doi.org/10.1111/cogs.13062

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6), 1176-1190. https://doi.org/10.1080/17470218.2013.850521

Wechsler, D. (2013). *Wechsler Preschool & Primary Scale of Intelligence - Fourth UK Edition (WPPSI – IV - UK)*. London: Pearson.

Whiteside, K. E., & Norbury, C. F. (2017). The Persistence and Functional Impact of English Language Difficulties Experienced by Children Learning English as an Additional Language and Monolingual Peers. *Journal of Speech, Language, and Hearing Research*, 60(7), 2014-2030. https://doi.org/10.1044/2017_jslhr-l-16-0318

Wiig, E. H., Secord, W. (1992). *Test of Word Knowledge*. U.S.A.: The Psychological Corporation.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: PCA details for Memory and Vocabulary Depth

Appendix S2: Backwards Words Recall Task

Appendix S3: Comprehension Monitoring task

Appendix S4: Listening comprehension task: Details of piloting and question selection

Appendix S5: Listening comprehension task: Stories

Appendix S6: Language questionnaire