A longitudinal study of theory of mind and listening comprehension: Is preschool theory of mind Important?

Jackson, S., Slade, L., Levy, J.P. and McCormick, S.F.

This is the accepted version of:

Sophie Jackson, Lance Slade, Joseph P. Levy, Samantha Felicity McCormick,

A longitudinal study of theory of mind and listening comprehension: Is preschool theory of mind important?,

Journal of Experimental Child Psychology,

Volume 219,

2022,

105388,

ISSN 0022-0965,

https://doi.org/10.1016/j.jecp.2022.105388.

A Longitudinal Study of Theory of Mind and Listening Comprehension: Is preschool Theory of Mind Important?

Sophie Jackson
University of Roehampton, Department of Psychology, Whitelands College, London SW15 4JD, UK
Present address: University of Derby, School of Psychology, Kedleston Road, Derby DE22 1GB
s.jackson@derby.ac.uk

Lance Slade
Canterbury Christ Church University, School of Psychology and Life Sciences, Canterbury, Kent, CT1 1QU, UK
lance.slade@canterbury.ac.uk

Joseph P Levy
University of Roehampton, Department of Psychology, Whitelands College, London SW15 4JD, UK
j.levy@roehampton.ac.uk

Samantha Felicity McCormick
Royal Holloway University of London, Department of Psychology, Egham, London, TW20 0EY, UK
Sam.McCormick@rhul.ac.uk

Corresponding author: Sophie Jackson (s.jackson@derby.ac.uk)

To cite this article: Jackson, S., Slade, L., Levy, J. P., & McCormick, S. F. (2022). A longitudinal study of theory of mind and listening comprehension: Is preschool theory of mind important? *Journal of Experimental Child Psychology*, *219*, 105388.

Abstract

Theory of mind has been shown to be important for listening comprehension for children at a range of ages. However, there is a lack of longitudinal evidence for a relationship between early theory of mind and later listening comprehension. The aim of this study was to examine if preschool theory of mind has a longitudinal direct effect on later listening comprehension over and above the effects of concurrent theory of mind. A total of 147 children were tested on measures of theory of mind, working memory, vocabulary and grammatical knowledge at Time 1 and Time 2 (T1 mean age= 4;1 years, T2 mean age = 5;11 years). Additionally, at Time 2, listening comprehension, comprehension monitoring and inference making measures were taken. Data was fitted to concurrent and longitudinal models of listening comprehension. Concurrent findings at Time 2 showed theory of mind to have a direct effect on listening comprehension. However, longitudinal findings showed that earlier theory of mind in preschool (Time 1) did not have a direct effect on listening comprehension 22 months later, instead there was only an indirect effect of earlier theory of mind on later listening comprehension via concurrent theory of mind (Time 2). Taken together, the results give further support for the importance of theory of mind for listening comprehension but show that there are limited additional benefits of early theory of mind acquisition. Implications for the development of children's listening comprehension are discussed.

Keywords: Theory of mind, Listening comprehension, Longitudinal Preschool
Direct effect Language comprehension

Theory of mind allows one to predict and explain others' behaviour based on the understanding of their mental states (Premack & Woodruff, 1978). It is largely accepted that children's general language ability is linked to their theory of mind understanding, as measured by false-belief tests. Indeed, Milligan, Dack and Astington's (2007) meta-analysis demonstrates a moderate to large effect size across age and a variety of language skills which holds across time. More recently, research has also suggested a link between concurrent theory of mind and listening comprehension, both in younger children aged 5-8 years (Kim, 2015, 2016, 2017, 2020a; Pelletier & Beatty, 2015) and older children aged 12-13 years (Ebert, 2020a), and in different languages and cultures including US English (Kim, 2017), German (Ebert, 2020a; Ebert, 2020b) and Korean (Kim, 2016). Yet contrary findings do exist, with Strasser and Rio (2014) finding that theory of mind did not make an independent contribution to listening comprehension in a sample of 4–6-year-olds. In addition to these mixed findings, there has been limited longitudinal research examining this relationship. In the current study we aimed to expand the understanding of theory of mind and listening comprehension by examining theory of mind as both a concurrent and early longitudinal predictor of listening comprehension in order to assess whether theory of mind ability in preschool impacts on later listening comprehension over and above the known effects of concurrent theory of mind ability. The aim was to determine if children's preschool theory of mind provides additional benefits for listening comprehension beyond its established concurrent effects at later ages.

Listening comprehension can be defined as using lexical information to understand discourse; more simply put it is the ability to understand what is heard (Hogan, Adolf &

Alonzo, 2014). It includes both the understanding of individual words, phrases and sentences, as well as larger elements of discourse. Theory of mind has been argued to contribute to children's listening comprehension ability by helping the listener to understand the intentions, desires and viewpoints of the speaker(s) (Kim & Pilcher, 2016), and allow better awareness of social information and details within the spoken passage (Dore, Amendum, Golinkoff, & Hirsh-Pasek, 2018). - Kim, et al. (2016) argue successful comprehension ultimately requires construction of a "situation model" or a "mental model" (Graesser, Singer, & Trabasso, 1994; Zwaan, Langston, & Graesser, 1995), that is, a mental representation, of what the passage is about (Kintsch, 1988). This mental representation may include information about characters, intentionality (or goals) and causation (Graesser, et al., 1994), and therefore theory of mind is a clearly an important source of information. Theory of mind may also, however, contribute more directly to the model building itself, in a domain general sense. Perner (1991) describes the possession of a theory of mind as mental model building in that a belief is a mental model of the world. Passing a false belief task shows that the child is able to understand (build) a model of both the world and how someone thinks about it. Clearly, this ability could assist directly with the creation of mental models during listening comprehension. The developmental trajectory of such a relationship is not yet clearly established, but it has been implicated in a number of studies investigating listening (and reading) comprehension (Atkinson, Slade, Powell, & Levy, 2017; Kim, 2017).

Much of the research into theory of mind and listening comprehension has come from Kim's cross-sectional data from five- to eight-year-olds in Korea and the USA (Kim, 2015, 2016, 2017, 2020a, 2020b). Using structural equation modelling, Kim proposes a hierarchical model of listening comprehension, in which lower-level skills predict high-level skills which in turn predict listening comprehension. The model includes the low-level skills of working memory, vocabulary, grammatical knowledge, and the high-level skills of comprehension monitoring and inference making, skills which are well evidenced as predictors of listening comprehension in a variety of studies (e.g. Alonzo, Yeomans-Maldonado, Murphy, & Bevens, 2016; Florit, Roch, Altoè, & Levorato, 2009; Florit, Roch, & Levorato, 2011; Strasser & Río, 2014). Importantly, Kim's model also includes theory of mind, which makes a significant direct contribution to listening comprehension when controlling for all other skills in the model (Kim, 2015, 2016, 2017, 2020a, 2020b). In later work (e.g. Kim, 2017), the model is named the Direct and Indirect Effects model of Text comprehension (DIET) as skills in the model are shown to have both direct paths to listening comprehension as well as indirect paths via other skills. This model provides, to date, the most comprehensive account of listening comprehension and its sub-skills which holds across early to middle childhood and across languages (Korean and US English).

However, there are limitations of the work supporting the DIET model. It is important to note that although the model refers to demonstrating effects, the concurrent and correlational nature of the studies means that the direction of relation in the model remains theoretical (Kim, 2015). Kim acknowledges that future work should track the development of listening comprehension and its sub-components across time to validate the model (Kim, 2015, p. 30). Recently, Kim (2020a) attempted to address this by exploring the stability over time of theory of mind to predict listening comprehension, by comparing the model for the same group of children when they were in Kindergarten (mean age = 5;4 years) to when they were in Grade 2 (mean age = 7;2 years). Findings showed that at both time points, after controlling for other variables in the model, theory of mind made a significant direct contribution to listening comprehension. Importantly, the effects (standardized regression weights) were similar in Kindergarten and Grade 2, indicating l stability over time. However, these findings only demonstrate that theory of mind is a concurrent predictor of listening comprehension i.e., that theory of mind has a concurrent effect on listening comprehension, and that this is consistent across developmental timeframes. The findings do not tell us anything about theory of mind as a longitudinal predictor i.e., whether earlier theory of mind development helps children to develop better future listening comprehension.

To date, only one researcher has examined the effects of earlier theory of mind on later listening comprehension (Ebert, 2020a; Ebert 2020b). Ebert (2020a) conducted longitudinal work in a study spanning ten years with German speaking children. The findings showed that earlier theory of mind at five years (mean = 5;6 years) had no effect, either directly or indirectly via later language ability, on later listening comprehension 7 years later, aged 12 (mean = 12;8 years). Instead, only children's advanced theory of mind at age 12 years (mean = 12;8 years) had a direct effect on listening comprehension 11 months later at age 13 (mean = 13;7). However, in a further longitudinal study, this time using the DIET model as a theoretical framework, Ebert (2020b) claimed that theory of mind measured at 5;6 years made a significant direct contribution to listening comprehension over 8 years later when children were aged 13;7. Yet, it is important to note here that this path was only marginally significant ($p < .10$). Given this inconsistent pattern of findings, further research is needed to explore these relationships. Additionally, Ebert (2020a) and Ebert (2020b) did not assess theory of mind at their first time points (mean age = 3;6 years) and so this research was unable to examine the effects of theory of mind in the preschool years on later listening comprehension, i.e. at 3-4 years - a critical time for theory of mind development (Perner, 1991; Wellman, Cross, & Watson, 2001). Examining early (preschool) theory of mind as a predictor of later listening comprehension was therefore a key aim of the current study.

Current study

The primary goal of this study was to examine the longitudinal relationship between theory of mind and listening comprehension by assessing theory of mind both concurrently and as a longitudinal predictor of listening comprehension. The current study sought to further develop recent work into the DIET model (Kim, 2020a) which has shown the relationship between theory of mind and listening comprehension to have stability over time, and recent work by Ebert (2020a; 2020b) which provides some (although mixed) evidence for theory of mind as a longitudinal predictor of listening comprehension. The current study expanded the work of Kim (2020a) and Ebert (2020a; 2020b) in three main ways. Firstly, it extended the DIET model to a UK population. Secondly, our study assessed theory of mind even earlier than Ebert's most recent study (2020b), when children were in preschool (aged 3 to 4 years of age) which is important because this is widely considered to be the most crucial age for the development of theory of mind (Perner, 1991; Wellman, et al., 2001;). Lastly, it looked both at theory of mind as a longitudinal temporally proximal, and, as a distal predictor of listening comprehension. This was done by including, within one model, a direct path from earlier theory of mind to later listening comprehension (a temporally distal predictor), and an indirect path from earlier theory of mind to later listening comprehension via concurrent theory of mind (a temporally proximal predictor). In addressing these points, the current study aimed to replicate the finding of a stable relationship between theory of mind and listening comprehension shown by Kim (2020a), but also to clarify if earlier theory of mind ability contributes additionally beyond this stability.

The DIET model proposes a direct effect of theory of mind on listening comprehension at each age tested above and beyond the effects of current language abilities, as measured by standardised language testing. Kim (2020a) demonstrates stability over time of this relationship at ages 5 and 7. This suggests a robust, direct effect of theory of mind on concurrent listening comprehension that reflects the benefits of improving theory of mind abilities at these ages. However, the source of this improvement is unclear; specifically, to what extent early (preschool) theory of mind ability is important. Theoretically, early theory of mind may indirectly contribute to later theory of mind ability and therefore indirectly benefit listening comprehension. Alternatively, there may be a direct contribution of early theory of mind over and above its concurrent effect on emerging listening comprehension i.e., there may be something important about early acquisition of theory of mind beyond this stability. Acquiring early theory of mind understanding may provide important additional experience of using theory of mind, for example when making sense of spoken passages, over and above the effects of later theory of mind understanding (see Atkinson, Slade, Powell and Levy, 2017, for a similar claim regarding the role preschool theory of mind in later reading comprehension). The DIET model as characterised by Kim et al. (2020a) does not distinguish between these two alternative accounts.

To reflect current thinking, the model structure used here was based on the most recent DIET model (Kim, 2020b; see Figure 1). First a cross-sectional (concurrent) model of listening comprehension when children were 5;11 years was tested (Time 2; Figure 1). Our aim, here, was to validate the DIET model for a UK population. Replicating the model in a UK sample is novel and provides important evidence for its robustness and generalizability (Duncan, Engel, Claessens, & Dowsett, 2014; Koole & Lakens, 2012). To our knowledge this was first study to fully replicate the model beyond the original research group as Ebert (2020b) tested only a section of the model. In line with previous work into the DIET model it was expected that the model would be a good fit and that theory of mind would make a direct contribution to listening comprehension.

Next, a longitudinal model was run (Figure 2) in which earlier skills at age 4;1 years (Time 1) were tested for their ability to predict listening comprehension 22 months later at age 5;11 years (Time 2). Importantly, this model included both a direct path from earlier theory of mind (Time 1) to later listening comprehension (Time 2), as well as an indirect path from earlier theory of mind to later listening comprehension via concurrent (Time 2) theory of mind. This was to test theory of mind as both a temporally proximal (i.e., concurrent) predictor and temporally distal (i.e., earlier) predictor of listening comprehension and would clarify whether earlier theory of mind ability contributes additionally to later listening comprehension over and above concurrent theory of mind. It was hypothesized that theory of mind would be both a temporally proximal and distal predictor of listening comprehension and therefore both these paths would be significant.

Material and methods

Participants

This study followed 147 (73 boys, 74 girls) UK children across 22 months from when they were in preschool (Time 1; mean age = 4;1 years, SD = 4.30, range = 3;2 - 4;9) to Year 1 (Time 2; mean age = 5;11 years, SD = 4.07, range = 5;1 – 6;8). Initially, 204 children were recruited but there was an attrition rate of 28% (57 children), primarily because children changed circumstances. This compares favourably with other, similar longitudinal research (e.g., 28% in Consortium & Chiu, 2018 and 29% in Lonigan et al., 2000). Children who left the project did not differ significantly from children who remained in the sample in terms of age and gender, nor vocabulary or non-verbal ability (as measured at Time 1). The only difference that should perhaps be noted is that children with English as an additional language (EAL) were significantly more likely to leave the study ($\chi(1) = 21.52$, $p <. 001$) with 62.5% of those who left having EAL. Only the 147 children with full data sets were used in the analysis.

Children were recruited from South East England and North West England in two cohorts; Cohort 1 (n = 107) was recruited directly via schools, and Cohort 2 (n = 40) was recruited directly via parents. Distributional properties and correlations of all variables were highly similar for the two cohorts and therefore their data was combined for all analyses. Children primarily came from well-educated families; 69% (of the 76% of mothers who completed an SES questionnaire) of mothers had at least an undergraduate degree. Some children had EAL (24.5%; 17 languages represented). This proportion of EAL children roughly represents the

UK mean figures suggesting a 20.1% rate of EAL in UK classrooms (Department for Education, 2016).

Measures

Children's theory of mind, vocabulary, knowledge of grammar and working memory were assessed at both time points. In addition, non-verbal ability was measured at Time 1 and listening comprehension, comprehension monitoring and inference making were assessed at Time 2. As a key aim of our research was to replicate and extend the DIET model, therefore where possible measures were selected to be in line with this previous work (e.g., Kim 2020a).

Theory of mind

At Time 1, two first-order false belief tasks, the unexpected contents task (Hogrefe, Wimmer, & Perner, 1986)) and the unexpected locations task (Wimmer & Perner, 1983), were administered to assess theory of mind, and at Time 2 the Strange Stories task (White, Hill, Happé, & Frith, 2009) was used. The use of different types of theory of mind measures at the two time points was to ensure that measures were age appropriate and is consistent with other longitudinal studies into theory of mind (e.g. Atkinson, et al., 2017; Devine, White, Ensor, & Hughes, 2016) including those examining listening comprehension (e.g. Ebert, 2020a).

For Time 1, in all analysis a composite of these two false belief tasks was used for the theory of mind variable. In the unexpected contents task children must recall their own false belief and attribute and explain the false belief of a character. Children were introduced to a character who was then taken out of sight while they were asked to predict the likely contents of a tube of “Smarties”. After the children had guessed, they were shown that the tube unexpectedly contained pencil crayons. The box was closed, and the children were asked again what was inside (control question). The control question needed to be answered correctly to gain credit on the test questions. The first test question required children to say what they thought was in the box before they looked inside (test question 1). The toy character was reintroduced with a reminder that the character had not seen inside the box. Children were then asked what the toy character would say was in the tube (test question 2) and why she would say this (test question 3; justification question). Children were awarded one point per correct test question (maximum score: 3). As the scoring of this measure required a judgment decision, 25% was scored by a second researcher and Cohen’s Kappa was calculated. Agreement was excellent with 96% of total agreement (Kappa = .94, Altmann, 1990).

The unexpected locations task requires children to predict and explain a character’s false belief about a transferred object. The children watched the experimenter narrating a story with small dolls where one character (Sally) was shown to play with a ball before placing it in a blue box and leaving the scene. Next another character (Anthony) was introduced and shown to go into the blue box, pick the ball up and play with it, and afterwards place the ball in the red box. On Sally’s return children were asked the test question requiring them to state where Sally would look for her ball first and also give a justification for their answer. This was followed by two control questions which needed to be answered correctly to gain credit for the test questions. Children were awarded one point for answering the test question correctly (blue box) and a further point for an appropriate justification if the test question had been answered correctly. The maximum score was two. Again, 25% was scored by a second researcher and agreement was excellent (98%, Kappa = .96). For Time 1, the theory of mind variable was created using a composite of the unexpected contents task and the unexpected locations task, therefore the overall maximum score for theory of mind at Time 1 was five.

At Time 2 five of the eight belief-based misunderstanding stories were used from White, et al. (2009). These were Stories 1 and 2 which involve a double-bluff, Story 3 which involves deception, and Stories 7 and 8 which involve misunderstandings. These stories, devised by White, et al., (2009), are based on the initial work by Happé (1994) and successful performance required children to attribute mental states to a story character to explain their behaviour. Each story included a sequence of three coloured cartoon pictures with an audio narration, presented on a tablet. After each story, the researcher asked a corresponding question to assess whether the child had understood the misunderstanding, or the untruth told. Although this task was originally designed for slightly older children (7-12 years) it has been used successfully with children of a similar age (5- and 6-year-olds) by other researchers, for example Kirk et al., (2015) and O'Hare, Bremner, Nash, Happé, & Pettigrew (2009). Responses were scored in line with White, et al. (2009) with no points awarded for irrelevant, or factually incorrect responses. One point was given for correct factual information but in the absence of full understanding of the lie or misunderstanding and its consequence. Two points were given if an answer showed an advanced understanding of the character's beliefs or motives. The total possible score was 10. Again, 25% were scored by a second researcher with excellent agreement between scoring (87%, Kappa = .83).

Vocabulary

At both time points receptive vocabulary was measured using the long form of the British Picture Vocabulary Scale: Third Edition (BPVS-III; Dunn & Styles, 2009). In line with standard protocol, children listened to a word spoken by the researcher and were asked to indicate the word's meaning by pointing to one of the four coloured pictures. At both time points the maximum score was 168.

Knowledge of grammar

Knowledge of grammar was measured using the sentence structure subset of the Clinical Evaluation of Language Fundamentals-*Preschool 2^{UK}* (CELF-Preschool 2; Semel, Wiig, & Secord, 2004) at Time 1 and the same subset of the Clinical Evaluation of Language Fundamentals-4^{UK} (CELF-4; Wiig, Secord, & Semel, 2006) at Time 2. In line with standard protocol, children listened to a spoken sentence and were asked to point to one of four coloured pictures which depicted that sentence. For Time 1 the maximum score was 22 and 26 at Time 2.

Working memory

At both time points the reverse word span task (Slade & Ruffman, 2005) was used to measure working memory. This is a modified version of the Backwards Digit Span task (Davis & Pratt, 1995). This oral task required children to reverse short sets of words spoken by the researcher. Following practice trials, at Time 1 the test trials included three sets of two words e.g. "*scarf – coat*", and three sets of three words e.g. "*plane – dog – pear*". At Time 2 there was an addition trial of three words. Children were awarded one point for correctly reversing two words, and two points for correctly reversing three words, and for trials of three words, a half point was given for reversing two words that were not adjacent. Therefore, the maximum score was nine at Time 1, and 11 at Time 2.

Inference making

At Time 2 only, inference making was measured using an oral story which required the ability to make both local coherence inferences and global coherence inferences. The story used was entitled "A new pet" and was developed by Consortium and Muijselaar (2018) based on research with older children (Cain & Oakhill, 2014; Cain & Oakhill, 1999). Children were required to answer the eight inference questions, following the story's narration by the researcher. Four questions involved local inferences and four required children to make global inferences. For example, children were asked the question "*What sort of animal was Sparky*?" and had to infer he was a dog from the knowledge that he was soft, furry and playful and had a kennel. The maximum total score was eight. Internal consistency for the present sample was α =.60.

Comprehension monitoring

At Time 2 only, comprehension monitoring was measured by children's ability to monitor if six oral stories made sense. The six stories (as used by Yeomans-Maldonado, 2017) were read aloud by the researcher. Following each story children judged whether the story *"made sense*" or "*did not make sense*". For example, one of the stories that did not make sense first stated that it was a girl's sixth birthday, and later that she had ten candles on her birthday cake. The maximum possible score was six. Three of the stories "made sense" and three "did not make sense". Internal consistency for the present sample was α =.73.

Non-verbal ability

Non-verbal ability was controlled for in all models. At Time 1 only, the block design subset of the Wechsler Preschool and Primary Scale of Intelligence - Third Edition (WPPSI-III; Wechsler, 2002) was used to assess non-verbal ability. Here children were required to recreate several geometric patterns of increasing difficulty using coloured blocks. The maximum score was 40.

Listening comprehension

At Time 2 listening comprehension was assessed using two measures, the Neale Analysis of Reading Ability-Second Revised British Edition Form 1 (NARA-II; Neale, 1999) and the listening comprehension subset of the Oral and Written Language Scales (OWLS-II; Carrow-Woolfolk, 2011). The two measures were significantly related (r = .62, p < .001). Using two measures of listening comprehension is in line with all other work into the DIET model and allowed us to capture all of what listening comprehension encompasses i.e., the understanding of both individual phrases and sentences (through the OWLS), as well as larger elements of discourse (through the NARA).

The NARA was originally designed to measure reading comprehension but has also been used as a measure of listening comprehension whereby stories are read aloud to participant by the researcher and children are asked questions to gauge their comprehension of the story (as used by Bowyer-Crane, et al., 2008; Cain & Bignell, 2014; Nation, Cocksey, Taylor, & Bishop, 2010). All passages used were fiction and told stories about the protagonist's actions. For example, a story about a girl called Kim, who was on her way to school when she saw two children on bikes crash into each other. Kim ran to help, only to find out from the children that they were taking part in a staged road safety lesson. The comprehension questions consisted of both literal questions (e.g. "Where was Kim going?") and inferential questions (e.g. "How do you think Kim felt?"). Scoring was similar to that used by Nation, et al., (2010) and ranged from 0-44.

The OWLS listening comprehension subset assesses comprehension of sentences and short passages and consists of 130 items, arranged in increasing order of difficulty. Each item was presented by reading the verbal stimulus aloud while the child looked at four coloured pictures numbered 1 to 4. An early example of an item is: *"Show me the girl saying, 'good-bye'",* to which the child had to select the meaning from: a picture of a girl putting her coat on, a girl hurrying along a path waving to a woman (correct answer), a boy getting into a car as he waves at a man, and a girl and woman planting flowers in a garden. A later example is:

"The boy to whom the girl with the broken arm had given a football brought her a glass of water", to which the child had to select the meaning from: a picture of a girl talking to a boy with a broken arm holding a ball, a picture of a boy giving a girl with a broken arm and a ball a glass of water, a picture of a boy with a ball and a girl with a broken arm talking, and a picture of a girl with a broken arm being given a glass of water by a boy with a ball (correct answer). The test was administered in line with standard protocol. The maximum total score was 130.

Procedure

At both time points children were tested individually in a quiet area either outside their classroom (Cohort 1) or within their home (Cohort 2). Children were initially tested in preschool (Time 1) and again 22 months later when they were in Year 1 (Time 2). Testing was part of a wider study whereby other measures were also administered. At each time point children either took part in two 20-minute sessions (Cohort 1) or one 40 -to- 60-minute session (Cohort 2). Within these sessions the administration of measures was fully counterbalanced.

Data analysis strategy

The primary data analysis strategy was structural equation modelling (SEM) using AMOS Version 26 (Arbuckle, 2016). Observed variables were used for all language and cognitive skills except listening comprehension where a latent variable was created consisting of two measures (NARA and OWLS). The use of this latent variable to measure listening comprehension was in line with Kim (2020a; 2020b) and all other previous work on the DIET model. Model fits were evaluated by using the chi-square statistic (χ^2), comparative fit index (CFI), Tucker-Lewis Index (TLI), and root mean square error of approximation (RMSEA). A non-significant chi-square, CFI and TLI values equal to or greater than .95 and RMSEA values below .08 indicate an excellent model fit (Kline, 2015). Typically, CFI and TLI values greater than .90 are considered to be good and below this poor, and RMSEA values greater than .10 indicate a poor fit (Kline, 2015). In all models age and Time 1 non-verbal ability was controlled for by entering these variables into the model first in line with the standard procedure of AMOS SEM (Arbuckle, 2016), however, models were also re-run without the control variables.

Results

Descriptive and preliminary analysis

Table 1 shows descriptive statistics including means, standard deviations, minimum values, maximum values, skewness and kurtosis for each variable. Correlations between measures are displayed in Table 2. As outlined in Table 1, there was some minimal missing data therefore, full information maximum likelihood estimation was used for all further analysis. All variables were normally distributed as indicated by skewness (± 1) and kurtosis values (±2), with the exception of working memory at Time 1. Pearson correlations were calculated for all variables and are reported in Table 2. However, for Time 1 working memory Spearman's rank was also calculated but as these were very similar to the Pearson correlation only Pearson is reported here.

Skewness and kurtosis met the normality assumption for SEM (e.g. skewness $< +3/-3$ and kurtosis values $<10/-10$; Kline, 2005) and therefore SEM models could be run. Correlations are shown in Table 2; theory of mind measures at Time 1 and Time 2 were weakly to moderately related to listening comprehension measures at Time 2 ($.28 \leq rs \leq .32$).

Table 1

Descriptive statistics

Variable	N	Min-Max	Mean (SD)	Skew	Kurtosis
T1 ToM	147	0-5	2.14 (1.68)	.23	-1.10
T1 BPVS	147	10-102	54 (17.43)	-.11	-.52
T1 CELF	146	0-21	11.99 (4.41)	-.63	-.09
T1 WM	146	0-8	1.09 (2.17)	1.88	2.17
T2 ToM (SS)	146	0-8	2.41 (1.53)	.87	1.06
T2 BPVS	147	48-130	88.03 (12.67)	.12	1.22
T2 CELF	147	7-26	20.71 (3.67)	-.93	1.01
T2 WM	147	0-11	7.01 (3.44)	-.42	-1.10
T2 Inference	145	0-8	4.29 (1.89)	-.22	-.65
T2 Comp mon	146	0-5	2.80 (.99)	.15	-.26
T2 NARA	146	1-16	6.80 (3.52)	.67	-.32
T2 OWLS	147	36-104	65.28 (13.26)	.56	-.16
T1 WPPSI	147	0-32	16.31 (5.99)	-.41	.25

Note. All scores are raw scores expect T1 ToM which is a composite of the two theory of mind measures. T1 = Time 1; T2 = Time 2; Min–Max = minimum–maximum; Skew = skewness; ToM = theory of mind; BPVS = British Picture Vocabulary Scale; CELF = Sentence Structure Subset of the Clinical Evaluation of Language Fundamentals; WM = reverse word span task measure of working memory; SS = Strange Stories; Inference = inference making stories; Comp mon = comprehension monitoring stories; NARA = Neale Analysis of Reading Ability; OWLS = Listening Comprehension subset of the Oral and Written Language Scales; WPPSI = Block Design subset of Wechsler Preschool & Primary Scale of Intelligence.

Table 2

Correlations between all variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. T1 ToM	-														
2. T1 BPVS	.36**	-													
3. T1 CELF	.45**	.65**	-												
4. T1 WM	.36**	.41**	.41**	-											
5. T2 ToM	.45**	.38**	.38**	.43**	-										
6. T2 BPVS	.43**	.59**	.57**	.35**	.35**	-									
7. T2 CELF	.26**	.46**	.38**	.19*	.28**	.47**	-								
8. T2 WM	.42**	.47**	.51**	.44**	.39**	.52**	.40**	-							
9. T2 Inference	.23**	.45**	.39**	.20*	.16	.37**	.20*	.22**	-						
10. T2 comp mon	.10	.20*	.22**	.15	.15	.19*	.13	.20*	.11	-					
11. T2 NARA	.32**	.39**	.36**	.46**	.29**	.32**	.23**	.41**	.43**	.20*	-				
12. T2 OWLS	.32**	.40**	.41**	.34**	.28**	.43**	.28**	.38**	.49**	.28**	.62**	-			
13. T1 WPPSI	.29**	.61**	.47**	.38**	.36**	.47**	.33**	.42**	.29**	.18*	.41**	.30**	-		
14. T1 Age	.08	.13	.16	.21*	.14	.13	.17*	.12	.03	-.04	.06	-.02	.15	-	
15. T2 Age	.001	.16	.13	.10	.18	.12	.23**	.15	.07	.002	.03	-.005	.14	.83**	-

Note. p <.05 = *, p <.001 = **. ToM = theory of mind; BPVS = British Picture Vocabulary Scale; CELF = Sentence Structure subset of the Clinical Evaluation of Language Fundamentals ; WM = reverse word span task measure of working memory; Inference = inference making stories; Comp mon = comprehension monitoring stories; NARA = Neale Analysis of Reading Ability; OWLS = Listening Comprehension subset of the Oral and Written Language Scales; WPPSI = Block Design subset of Wechsler Preschool & Primary Scale of Intelligence.

Main analysis

First a cross-sectional model of listening comprehension when children were 5;11 years (Time 2) was tested (Figure 1). Next a longitudinal model was run (Figure 2) in which earlier skills aged 4;1 month (Time 1) were tested for their ability to predict listening comprehension aged 5;11 years (Time 2). This model included both a direct path from earlier theory of mind (Time 1) to later listening comprehension (Time 2; distal path), as well as an indirect path (proximal path) from earlier theory of mind to later listening comprehension via concurrent (Time 2) theory of mind.

Cross-sectional analysis

The cross-sectional model is shown in Figure 1. Multivariate normality assumptions of variables in the model were checked and were not violated. The model had an excellent fit: $\chi^2(7) = 9.11$, $p = .25$; CFI = .99, TLI = .96, RMSEA = .06. Standardized path coefficients are shown in Figure 1 and all control for concurrent age and non-verbal ability (WPPSI scores). Importantly, theory of mind ($\beta = .24$, $p = .01$) was independently related to listening comprehension.

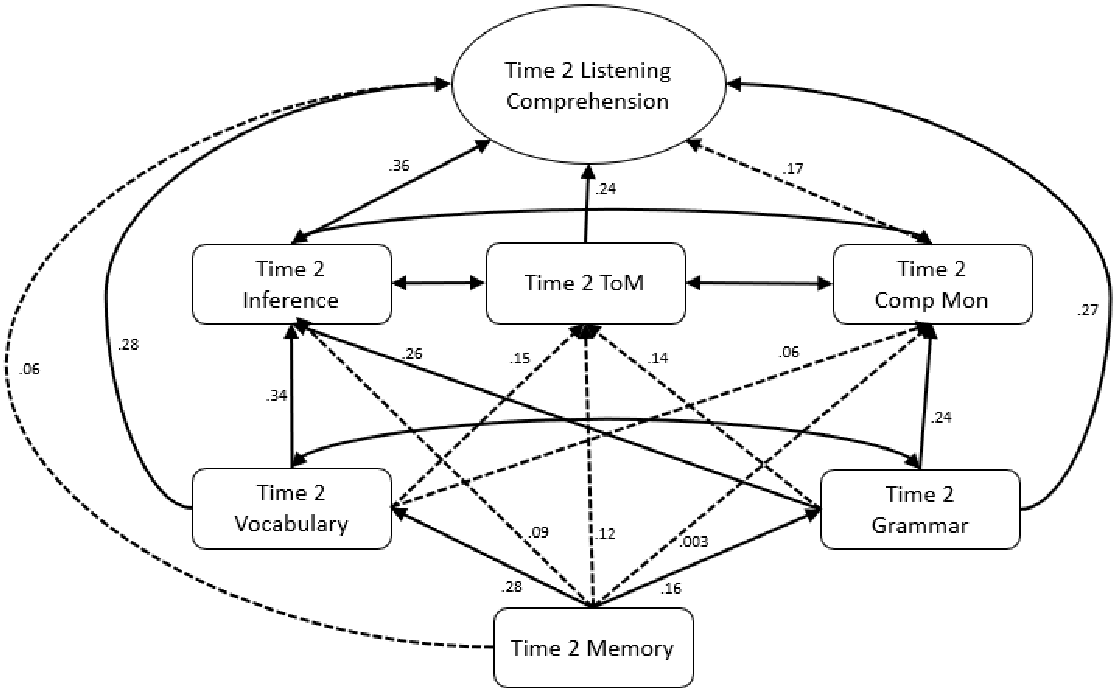

Fig 1: Cross-sectional model for predicting Time 2 listening comprehension using Time 2 skills after controlling for Time 2 age and Time 1 non-verbal ability. Solid paths are statistically significant at $p < .05$; dashed paths are not statistically significant. Two-sided arrows represent covariances. Memory = working memory, Grammar = grammatical knowledge, Comp Mon = comprehension monitoring, ToM = theory of mind, Inference = inference making.

Longitudinal analysis

The longitudinal model is shown in Figure 2. Multivariate normality assumption of variables in the model were checked and were not violated. The model fit had a good to excellent fit: $\chi^2(13) = 16.36$, $p = .23$; CFI = .99, TLI = .97, RMSEA = .04. Standardized path coefficients are shown in Figure 2 and all control for Time 1 age and non-verbal ability (WPPSI score). Importantly, Time 1 theory of mind was not independently related to Time 2 listening comprehension ($\beta = .09$, $p = .35$), but Time 2 theory of mind was ($\beta = .17$, $p < .05$). Moreover, Time 1 ToM and Time 2 ToM were independently related ($\beta = .23$, $p < .05$).

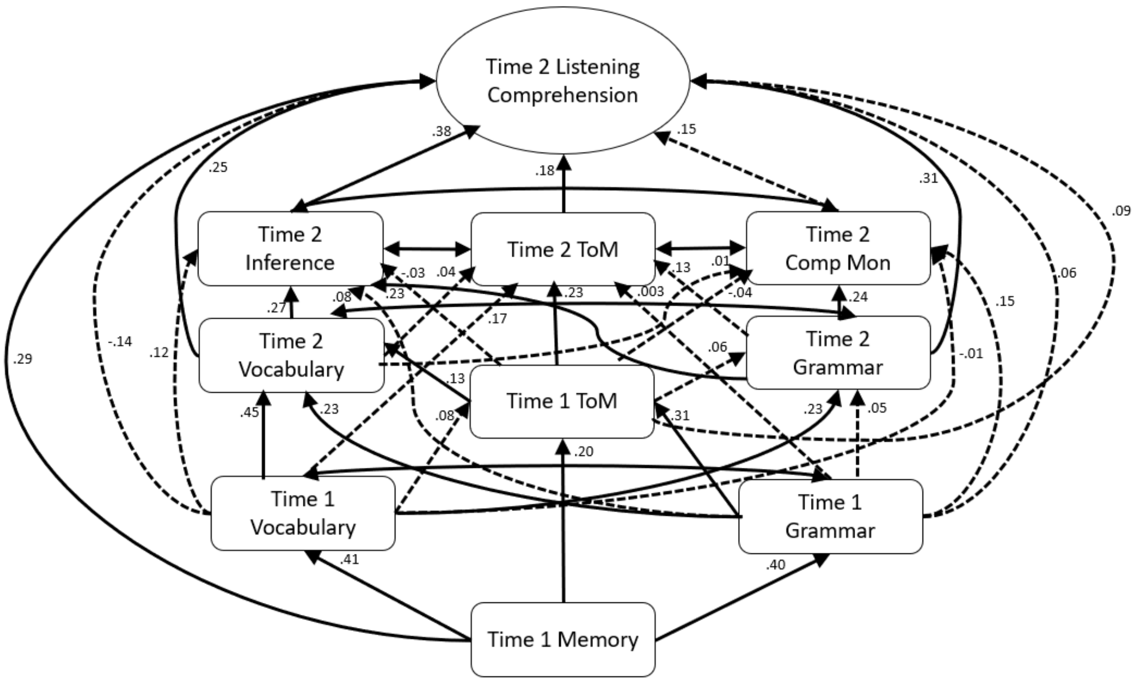

Fig 2: Longitudinal model for predicting Time 2 listening comprehension using Time 1 skills after controlling for Time 1 age and non-verbal ability. Solid paths are statistically significant at $p< .05$; dashed paths are not statistically significant. Two-sided arrows represent covariances. Memory = working memory, Grammar = grammatical knowledge, Comp Mon = comprehension monitoring, ToM = theory of mind, Inference = inference making.

Further analysis

Age and non-verbal ability were controlled for in both models. This was because, non-verbal ability consistently correlated with other variables, and age is a common control variable. However, as a check the two models were also run without controlling for age and non-verbal ability. When this was done the model fits did not change (i.e. the fit for the cross-sectional model was still excellent: $\chi2(5) = 4.75$ $p = .45$; CFI = 1.00, TLI = .1.05, RMSEA < .001. , and the longitudinal model was still a good to excellent fit: $\chi2\ (9) = 12.17$, $p = .20$; CFI = .99, TLI = .96, RMSEA = .05). Standardized path coefficients also remained very similar, for example in the cross-sectional model theory of mind and listening comprehension were still independently related (β = .27, $p = .01$), and in the longitudinal model Time 1 theory of mind was still not independently related to Time 2 listening comprehension (β = .12, $p = .27$).

In order to check there were no differences between the EAL children and their monolingual peers, multi-group analysis was used in AMOS to compare the fits of models for EAL children to English only speaking children. Neither models were found to fit significantly differently for these two groups.

Discussion

In the current study we aimed to gain further insight into the longitudinal relationship between theory of mind and listening comprehension in young children. Specifically, using a direct and indirect model we assessed theory of mind as a longitudinal or concurrent predictor of listening comprehension. This study builds on the recent work of Kim (2020a) and Ebert (2020a; 2020b) as it aimed to replicate the stability of the relationship between theory of mind and listening comprehension shown by Kim (2020a), and also to clarify if earlier theory of mind ability contributes additionally beyond this stability. We used the DIET model (e.g. Kim, 2017, 2020b) as a theoretical framework as this is the most comprehensive (and robustly tested) model of listening comprehension and its subskills, and importantly is the first model to include theory of mind.

Validation of the DIET model for a *UK population*

Concurrent findings (see Figure 1) supported the DIET model (Kim, 2020a) as our model was shown to have an excellent fit for a UK population at 5;11 years (Year 1). This is notable because until now the DIET model has not been fully tested outside of this research group and beyond US and Korean populations. This reinforces the DIET model as the most comprehensive and robustly tested model of listening comprehension and its subskills. This is important as Duncan, et al. (2014) stress the need for replication studies to provide evidence for robustness and generalizability in developmental psychology.

Crucially, in our model the significant path shown in the DIET model from theory of mind to listening comprehension (e.g. Kim, 2017, 2020a) was replicated. This strengthens the claim that theory of mind is important for listening comprehension (Ebert, 2020a; Kim, 2015, 2016, 2017, 2020a, 2020b; Pelletier, et al., 2015). Theory of mind perhaps aids listening comprehension because it can help the listener with the social information within a passage (Kim, et al., 2016). Additionally, both skills are concerned with the building of mental models (Kintsch, 1988; Perner, 1991), and therefore children who are adept at building mental models to understand the social world can use these skills when building a mental model of a passage of text during listening comprehension. In our study, the outcome measures of listening comprehension (NARA and OWLS) both required children to comprehend social information of characters from the passages, including building increasingly more complex mental models to understanding who did what and why. Our findings therefore support this theoretical suggestion of how theory of mind aids comprehension.

Beyond theory of mind, most paths within the model were also replicated. For example, both the significant path from inference making to listening comprehension, and the non-significant path from comprehension monitoring to listening comprehension were consistent with Kim (2020a). This provides support for the DIET model as a whole and shows that it can be generalised beyond the US and Korea to a UK younger population.

Theory of mind as a longitudinal predictor of listening comprehension

The longitudinal analysis aimed to test theory of mind as both a temporally proximal (concurrent) and a temporally distal (longitudinal) predictor of listening comprehension. In order to do this a model was run (Figure 2) which included both a direct path from earlier theory of mind (Time 1) to later listening comprehension (Time 2), as well as an indirect path from earlier theory of mind (Time 1) to later listening comprehension (Time 2) via concurrent theory of mind (Time 2). It was hypothesized that theory of mind is both a temporally proximal and distal predictor of listening comprehension and therefore both the indirect (proximal) and direct (distal) path would be significant. Assessing theory of mind as a longitudinal distal predictor extends the work of Kim (2020a) who confirmed only stability over time in the relationship between theory of mind and listening comprehension. If theory of mind is a direct longitudinal predictor of listening comprehension, then improving children's earlier theory of mind will directly help with their later listening comprehension.

The hypotheses were only partially supported. There was no evidence for theory of mind as a temporally distal predictor of listening comprehension because the direct path from earlier theory of mind (Time 1) to later listening comprehension (Time 2) was non-significant. There was, however, evidence for theory of mind as a temporally proximal predictor of listening comprehension as the indirect path from earlier theory of mind (Time 1) to later listening comprehension (Time 2) via concurrent theory of mind (Time 2) was significant. In other words, the path from Time 1 theory of mind to Time 2 theory of mind was significant, and the path from Time 2 theory of mind to Time 2 listening comprehension was significant. These findings suggest that although early theory of mind indirectly affects later listening comprehension via later theory of mind, improving children's earlier theory of mind will not help with their later listening comprehension *directly*. In general, *concurrent* theory of mind remains important for listening comprehension.

It was hoped that the findings would both support and extend the recent work of Kim (2020a) who found there to be stability between kindergarten and Grade 2 in the effects of theory of mind on listening comprehension, but who did not explore theory of mind as a longitudinal predictor of listening comprehension. The findings do support Kim (2020a) in that they show evidence for the stability of the relationship between theory of mind and listening comprehension because they provide evidence for the relationship at 5;11 years. The current study also goes beyond that of Kim (2020a) as it provides evidence that earlier theory of mind does not directly predict listening comprehension. Thus, earlier theory of mind ability does not seem to contribute additionally beyond this stability.

The findings also have a bearing on Ebert's work (2020a; 2020b). In one study, Ebert (2020a) found that earlier theory of mind at 5;6 years had no direct effect on later listening comprehension at the age of 12;8 years. However, in another study, Ebert (2020b) found that theory of mind measured at 5;6 years did make a (marginally) significant ($p < .10$) direct contribution to listening comprehension 8 years later. It should be noted however, that Ebert (2020b) only tested a section of the DIET model leaving out measures of inference making and comprehension monitoring. It is unclear therefore whether ToM would still make a direct contribution to later listening comprehension over and above these other higher order skills. Our finding of no direct relationship between earlier ToM (at 4;1 years) and later listening comprehension at (5;11 years) is clearly consistent with the lack of a direct link shown in Ebert (2020a). These consistent findings, taken across the different age ranges, suggests that theory of mind is not a direct longitudinal predictor of listening comprehension. It is worth noting however that Ebert (2020a) did show a direct longitudinal effect with older children across a short time period (11 months). Given our finding for a clear concurrent effect of ToM on listening comprehension, it is possible that 11 months, in this older age group, is more akin to a concurrent effect shown at a younger age which is a period of more rapid developmental change. To explore this future work should look at different longitudinal timeframe lengths.

Our findings suggest that theory of mind is a concurrent but not a direct longitudinal predictor of listening comprehension, rather, preschool theory of mind, predicts later theory of mind which in turn, predicts concurrent listening comprehension at the second time point. Theoretically a direct effect of Time 1 theory of mind on later listening comprehension is plausible. Early acquisition may provide important additional experiences in using theory of mind in making sense of spoken passages (e.g., Atkinson et al., 2017), which make a direct contribution to later listening comprehension over and above contributions of later theory of mind. However, this was not supported in the current study.

Limitations and further directions

It is possible that the lack of an earlier direct effect reflects different measures of theory of mind used at the different time points. Measures were chosen to be age-appropriate and are well used at these ages (e.g. Atkinson, et al., 2017; Devine et al, 2016), however it is plausible that these different types of theory of mind tasks are not measuring theory of mind in the same way (White, et al., 2009). Although not explicitly stated, the DIET model is assumed to view theory of mind as a broadly unitary construct which is tapped in comparable ways by measures at earlier and later ages. However, research presents mixed findings over whether false belief tests and more advanced measures of theory of mind measure the same underlying ability (Brent, Rios, Happé, & Charman, 2004; Mitroff, Sobel, & Gopnik, 2006), with some arguing that the underlying abilities assessed by these measures form, at best, a variety of loosely related social-cognitive skills (Hayward & Homer (2017). Therefore, whilst the Strange Stories were selected to tap the representation of belief-based misunderstandings, it can be argued that this task draws on more advanced and complex understanding of mental states beyond simple understanding of belief. These more advanced tasks are also more linguistically challenging, perhaps in ways that fail to be captured by the standardised language measures administered in this study and overlap somewhat with the demands of the outcome measure (listening comprehension). Taken together, these may explain the differences across longitudinal timeframes in the results of

Ebert (2020a) and the current study. More longitudinal studies are needed to further examine these issues.

In addition to the limitation discussed above, there are several other points to consider, many of which concern the measures used by the study. The outcome measures of listening comprehension (NARA and OWLS) were chosen to reflect previous work into the DIET model (e.g. Kim, 2020a) as a key aim of our study was to replicate and extend this model. Similarly, the skills of inference making and comprehension monitoring were included in our study to reflect the original DIET model, and measures were chosen accordingly. However, a limitation of both this study and the DIET model is that measures of inference making and comprehension monitoring can be considered similar to the measures of listening comprehension. For example, the inference making measure used here required children to listen to an oral passage and answer inference-based questions, and both the NARA and OWLS measures of listening comprehension at times required children to make inferences in order to answer correctly. Therefore, although beyond the remit of this current study, future research may wish to replicate the DIET model but with an effort to ensuring there is less overlap between measures.

All skills in the models were measured by observed variables (i.e. a single measure) but to reduce measurement error it would have perhaps been better to use latent variables (Russell, Kahn, Spoth, & Altmaier, 1998) i.e. several measures per skill. This said, although this would have been better from a measurement perspective it would have meant administering more measures to children at each time point which could have led to fatigue and therefore may have affected reliability of measures, particular at Time 1 when children were very young.

Future research should aim to assess reading comprehension in the same way that listening comprehension has been explored here. In parallel with models of listening comprehension the DIET research team additionally explore a similar model of reading comprehension which also includes a place for theory of mind (e.g. Kim, 2015, 2020a, 2020b), but as of yet this has not been explored longitudinally. Ebert (2020b) do begin to explore this but as with their listening comprehension analysis, only a section of the original model is tested. It is of crucial importance to continue to develop our understanding of the subskills of listening and reading comprehension and their early longitudinal predictors (Hjetland, Brinchmann, Scherer, Hulme, & Melby-Lervåg, 2020) given the social and educational benefits of strengthening early reading (Florit & Cain, 2011; Massonnié, Bianco, Lima, & Bressoux, 2019).

Conclusions

Overall, this study improves our understanding of the relationship between theory of mind and listening comprehension, specifically building on recent work by Kim (2020a) and Ebert (2020a; 2020b). The study validates the DIET model to a UK population, demonstrating that within this model concurrent theory of mind significantly and directly relates to listening comprehension. This research also replicates the stability across time of the relationship between theory of mind and listening comprehension presented by Kim (2020a). However, it suggests that earlier theory of mind ability does not contribute additionally beyond this stability. That is, it presents evidence that earlier theory of mind in preschool does not have a direct effect on listening comprehension 22 months later. Helping children's theory of mind at any age, is likely to be beneficial for their concurrent listening comprehension, but early work specifically on this ability does not necessarily provide additional benefits later in their developmental trajectory.

References

Alonzo, C. N., Yeomans-Maldonado, G., Murphy, K. A., & Bevens, B. (2016). Predicting second grade listening comprehension using prekindergarten measures. *Topics in Language Disorders*, 36, 312-333.

Arbuckle, J. (2016). IBM® SPSS® Amos™ 24 user's guide.: Amos Development Corporation

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.

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., Götz, K., & Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry*, 49, 422-432.

Brent, E., Rios, P., Happé, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced ToM tasks. *Autism: the International Journal of Research and Practice*, 8 (3), 283–299.

Cain, K., & Bignell, S. (2014). Reading and listening comprehension and their relation to inattention and hyperactivity. *British Journal of Educational Psychology*, 84, 108-124.

Carrow-Woolfolk, E. (2011). OWLS-II: Oral and Written Language Scales: Western Psychological Services.

Consortium, L. a. R. R., & Chiu, Y.-D. (2018). The simple view of reading across development: Prediction of grade 3 reading comprehension from prekindergarten skills. *Remedial and Special Education*, 39, 289-303.

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology*, 47, 25-31.

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology*, 52, 758.

Dore, R. A., Amendum, S. J., Golinkoff, R. M., & Hirsh-Pasek, K. (2018). Theory of mind: A hidden factor in reading comprehension? *Educational Psychology Review*, 30, 1067-1089.

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology*, 50, 2417.

Dunn, L., & Styles, B. (2009). BPVS III. British Picture Vocabulary Scale. Third Edition. London: GL Assessment.

Ebert, S. (2020a). Theory of mind, language, and reading: Developmental relations from early childhood to early adolescence. *Journal of Experimental Child Psychology*, 191, 104739.

Ebert, S. (2020b). Early language competencies and advanced measures of mental state understanding are differently related to listening and reading comprehension in early adolescence. *Frontiers in Psychology*, 11, 952.

Education, D. f. (2016). Statistical first release: Schools, pupils, and their characteristics. In D. f. Education (Ed.). London Department for Education.

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, 23, 553-576.

Florit, E., Roch, M., Altoè, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology*, 27, 935-951.

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, 119-138.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371.

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24, 129-154.

Hayward, E. O., & Homer, B. D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology*, 35(3), 454-462.

Hjetland, H. N., Brinchmann, E. I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review*, 100323.

Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech-language Pathology*, 16(3), 199-207.

Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, 567-582.

Kim, Y.-S. G. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, 86, 128-144.

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.

Kim, Y.-S. G. (2017). Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading*, 21, 310-333.

Kim, Y.-S. G. (2020a). Theory of mind mediates the relations of language and domain-general cognitions to discourse comprehension. *Journal of Experimental Child Psychology*, 194, 104813.

Kim, Y.-S. G. (2020b). Toward Integrative Reading Science: The Direct and Indirect Effects Model of Reading. *Journal of Learning Disabilities*. 53(6), 469-491.

Kim, Y.-S. G., & Pilcher, H. (2016). What is listening comprehension and what does it take to improve listening comprehension? In Interventions in Learning Disabilities (pp. 159-173): Springer.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163.

Kirk, E., Pine, K., Wheatley, L., Howlett, N., Schulz, J., & Fletcher, B. (2015). A longitudinal investigation of the relationship between maternal mind-mindedness and theory of mind. *British Journal of Developmental Psychology*, 33(4), 434-445.

Kline, R. B. (2015). Principles and practice of structural equation modeling: Guilford publications.

Koole, S. L., & Lakens, D. (2012). Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science*, 7, 608-614.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology*, 36(5), 596.

Massonnié, J., Bianco, M., Lima, L., & Bressoux, P. (2019). Longitudinal predictors of reading comprehension in French at first grade: Unpacking the oral comprehension component of the simple view. *Learning and Instruction*, 60, 166-179.

Mitroff, S. R., Sobel, D. M., & Gopnik, A. (2006). Reversing how to think about ambiguous figure reversals: Spontaneous alternating by uninformed observers. *Perception*, 35, 709–715.

Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry*, 51, 1031-1039.

Neale, M. D. (1999). Neale analysis of reading ability: Reader: ERIC.

O'Hare, A. E., Bremner, L., Nash, M., Happé, F., & Pettigrew, L. M. (2009). A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *Journal of Autism and Developmental Disorders*, 39(6), 916-928.

Pelletier, J., & Beatty, R. (2015). Children's understanding of Aesop's fables: relations to reading comprehension and theory of mind. *Frontiers in Psychology*, 6, 1448.

Perner, J. (1991). Understanding the representational mind: The MIT Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515-526.

Russell, D. W., Kahn, J. H., Spoth, R., & Altmaier, E. M. (1998). Analyzing data from experimental studies: A latent variable structural equation modeling approach. *Journal of Counseling Psychology,* 45, 18.

Semel, E., Wiig, E. H., & Secord, W. A. (2004). Clinical evaluation of language fundamentals-preschool-2 (CELF-Preschool-2). Psychological Corp.: San Antonio, TX, USA.

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, 23, 117-141.

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, 169-187.

Wechsler, D. (2002). WPPSI-III administration and scoring manual: Psychological Corporation.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655-684.

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development*, 80, 1097-1117.

Wiig, E., Secord, W., & Semel, E. (2006). CELF 4: Clinical Evaluation of Language Fundamentals Pearson.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.

Yeomans-Maldonado, G. (2017). Development of comprehension monitoring in beginner readers. *Reading and Writing*, 30, 2039-2067.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6, 292-297.