GLOSSA psycholinguistics

Pagmar, D., Arvidsson, C., Gerholm, T. N., & Uddén, J. (2023). Conversations between ages five and seven – Connectionsto executive functions and implicature comprehension. *Glossa Psycholinguistics*, 2(1): 17, pp. 1–41. DOI: https://doi.org/10.5070/G6011132

eScholarship University of California

Conversations between ages five and seven – Connections to executive functions and implicature comprehension

David Pagmar, Department of Linguistics, Stockholm University, Sweden, david.pagmar@ling.su.se

Caroline Arvidsson, Department of Linguistics, Stockholm University, Sweden, caroline.arvidsson@ling.su.se

Tove Nilsson Gerholm, Department of Linguistics, Stockholm University, Sweden, tove@ling.su.se

Julia Uddén, Department of Linguistics, Stockholm University, Sweden; Department of Psychology, Stockholm University, Sweden, julia.udden@ling.su.se

A language user must rely on several different abilities to carry out a conversation, e.g. the ability to acknowledge the conversational contributions of others, to respond appropriately, to stay on topic, etc. There are many aspects of the development of conversational conduct that are yet unknown. In this study, the longitudinal development of conversational conduct, as in acknowledging one's interlocutor's previous turn, was traced from age 5;0 to 7;2. We also investigated whether conversational conduct was predicted by core language skill, executive functions, and specific pragmatic abilities. Previous findings of productive morphosyntactic accuracy were replicated, while findings concerning longitudinal receptive vocabulary were not. We also found connections between childrens' conversational responses and executive functions, working memory, and the comprehension of conversational implicatures. The results suggest that conversational conduct is dependent on inferring communicative intentions, as well as being able to keep track of others' contributions and how they relate to previous turns.

1. Introduction

There are few other situations in which a language user displays her linguistic and extra-linguistic abilities as much as in face-to-face conversation. During first language acquisition, a child must learn to take several components into consideration when engaging in conversation, such as acknowledging what others have said, staying on the conversational topic, as well as furthering it. In the current study, the aim is to understand to what degree fundamental conversational behaviours, like acknowledging the turns of others, develop during middle childhood, and to what degree this development relates to other developing abilities, such as core language ability and executive functions. The current study is a first step towards answering questions concerning what is more important during the development of conversational conduct: inferential processes and social aspects of the interaction, or abilities that allow for structural aspects of interaction to be facilitated.

When talking about conversation, we refer to the interactional conduct where two or more language users take turns in a cooperative fashion (Grice, 1975) to send and receive communicative linguistic messages. A key aspect of carrying out a conversation is to acknowledge what one's interlocutor has said. This acknowledgement is part of the conduct referred to as *conversational contingency*. The notion of conversational contingency stems from a paper from Bloom, Rocissano, and Wootten (1982), in which they define *contingent speech* as utterances that (i) share the topic of the preceding utterance and, (ii) add information to it (1976: 528). Bloom et al.'s definition has since been adopted in studies examining naturalistic conversations in both typically and atypically developing children (Capps et al., 1998; Hale & Tager-Flusberg, 2005; Hoff-Ginsberg, 1998; Nadig et al., 2010; Tager-Flusberg & Anderson, 1991). Previous results show that adult conversationalists excel at acknowledging their interlocutor's previous turns, by following both or simply the first of Bloom et al.'s criteria. When examining the percentage of turns in which adults acknowledge their partner's previous turn, 92% (Pagmar, 2023). Children, however, do not display the same conversational behaviour. When comparing percentages, Swedish five-year-olds acknowledged their interlocutor's previous turn far less than adults (51% of total number of responses) (Pagmar et al., 2022). We know that children get better as communicators during the later preschool years (e.g. Wanska & Bedrosian, 1985), but we do not yet know the developmental trajectory of acknowledging one's interlocutor's previous turn.

There are several different components that can be taken into consideration when using the term *conversational conduct*. In the current study, conversational conduct is used in reference to the conversational outcome measures (presented below), in particular the fundamental conversational behaviour of acknowledging one's interlocutor's previous turn.

Conversational conduct is likely dependent on a number of abilities. Aspects of conversational ability have been connected to receptive vocabulary, working memory, SES, etc. In the current study, we investigate three potentially important skills: core language, executive functions, and specific pragmatic abilities: 1. **Core language**: Previous findings (Pagmar et al., 2022) suggest

that children's appropriate responses at the age of 5;0 were predicted by receptive vocabulary at the age of 4;0, as measured by the PPVT-4 (Dunn & Dunn, 2012). However, in the same study, concurrent measures of morphosyntactic accuracy did not show any effect on conversational conduct. 2. **Executive functions**: Although the evidence for a relationship between conversational conduct and executive functions is sparse, findings from Blain-Brière, Bouchard, and Bigras (2014) showed the number of children's utterances during semi-naturalistic conversation was negatively correlated with measures of inhibition and their willingness to respond to their interlocutor was positively associated with working memory. 3. **Specific pragmatic abilities**: Conversation is the conduct in which specific pragmatic abilities are displayed the most. Conversationalists need to make inferences, keep track of referents, follow social norms, etc. However, recent findings on children, aged 7;0–12;0, suggest no strong relationship between different types of specific pragmatic abilities, such as Implicature Comprehension, Local Textual Inference, and Pragmatic Violations (Wilson & Bishop, 2022). It is therefore of interest to examine how different pragmatic abilities, such as comprehension of conversational implicatures, referential production, etc., relate to conversational conduct during development.

In the present study, we examined the development of specific types of conversational behaviours in semi-naturalistic conversations. We made additional observations with the same participants as in Pagmar et al. (2022), two years after the first observations. This allowed us to compare individual children's conversational conduct and how it had changed from age 5;0 to 7;2. In addition to the conversational data, we obtained measurements of core language ability, executive functions, as well as a set of specific pragmatic abilities: comprehension of conversational implicatures, referential production, audience design, and narrative skill.

We entertain three potential outcomes concerning how the conversational conduct related to the other abilities that we were taking into consideration: (O1) To follow a conversational topic, a conversationalist must make sense of the content of their partner's contributions. If A struggles with making inferences about B's intentions, then A will have a hard time trying to provide contingent and appropriate responses. In this sense, we can at least think of parts of conversational conduct as a social pragmatic ability (Andrés-Roqueta & Katsos, 2017). If this is the case, then we can expect to see conversational conduct being explained by measures of pragmatic ability, specifically, comprehension of conversational implicatures. The included measures that test inferential processes of different levels (specifically, implicatures, audience design, referential production and narrative skill) are expected to agree with this suggestion. It is also possible that (O2) the conduct of acknowledging one's interlocutor's previous turn relies more heavily upon abilities such as executive functions, e.g. working memory. If that is so, then conversational conduct is perhaps not as dependent on pragmatic inferences. A can infer what B's intentions are, but mainly needs the ability to keep track of communicative content to be a good conversational partner. The included measures that test executive functions (the Wisconsin

Card Sorting Task and Backward Digit Span) are expected to agree with this suggestion. Another possibility is (O3): pragmatic abilities and executive functions are both important and work simultaneously, together affecting conversational behaviour, as well as each other.

In the literature, core language skill is the factor that has been shown to correlate with different aspects of children's pragmatic competence (Bernard & Deleau, 2007; De Rosnay et al., 2014). The inclusion of core language measures is therefore key when determining which factors aid conversational conduct specifically.

Throughout the paper, we will follow the same presentation structure. with four overarching headlines: 1. development of conversational conduct, 2. conversation and core language, 3. conversation and executive functions, and lastly, 4. conversation and specific pragmatic abilities.

1.1 Conversation during the later preschool years

Engaging in conversation is essential for making and maintaining friendships (see e.g. Hazen & Black, 1989) as well as when collaborating on problem-solving activities both in school and in the workplace. It is by no means a given that all measures of conversational and communicative skill will develop from the later preschool years, up to school start ($\approx$ 5;0–7;0).[1] In this section, we present studies showing contrasting results on conversational and communicative development during this developmental period.

Yliherva et al. (2009) investigated Finnish children's conversational ability, using the *Children's Communication Checklist* (Bishop, 1998). The results showed that Finnish four-year-olds differed from six-year-olds on two of the checklist's subscales: the Syntax subscale (grammar ability) and the Coherence subscale (to represent past and future events). Five-year-olds differed from the six-year-olds on the ability to use non-verbal communication such as eye contact, facial expression, gestures, etc. On the other hand, the subscale Use of Context (to comprehend and express utterances in different contexts in a relevant way) did not show a significant difference between the ages four and six.

There are other aspects of linguistic communication that remain unchanged from age four to six. Turkish children's use of demonstratives during age-paired dyadic interaction showed no reliable change between 4;0 and 6;0, but both ages differed in demonstrative distributions when compared to adult conversation (Küntay & Özyürek, 2006).

There are several studies on communicative development below the age of 5;0: conversational repair skills between 1;0 and 3;10 show that children have an emerging recognition of what is required of them by others during conversation (Forrester & Cherington, 2009); quality of preschool and specifically measures of teachers' instructional interactions and teachers' emotional interactions predicted aspects of both social and language development at 4;0 (Mashburn et al., 2008); parental

[1] In Sweden, children start first grade in the year that they turn seven.

input influences and guides pragmatic behaviours, for example indirectness, for children between 2;11–4;3 (Becker, 1994). These three findings alone exemplify that there are several factors, both internal and external to the child, that affect the development of language use.

In the current study, we were concerned with children's development as linguistic communicators, specifically, the conduct of acknowledging one's interlocutor's previous turn in conversation. Based on prior research, we can expect that syntactic ability, logically structured accounts of events, and gestural abilities are improving between 5;0 and 7;2, but we do not know how conversational conduct in itself is developing during the same period. We will offer an overview of how the conversational ability can be connected to core language ability, executive functions, and other specific pragmatic abilities.

1.1.1 Conversational conduct in the current study

We will now present how conversational conduct is represented in the current study.

We investigated the development of specific conversational behaviours from 5;0 to 7;2. The conversational responses of our participants were categorised in accordance with a coding scheme (Abbot-Smith et al., 2021; Pagmar et al., 2022), tracking conversational contingency in four types of responses:

I. to add information and further the topic
II. to acknowledge what was previously said (whether it furthers the topic or not)
III. to not acknowledge what was previously said
IV. to not respond at all

In the current study, percentages of types II-IV were held as the outcome measure, since these types directly track the conversational conduct of acknowledging the other's turn. Previous results have presented proportions of different types of conversational responses from children at 5;0: *appropriate responses* (corresponding to type II, considered a positive measure) mean = 51%, sd = 17%; *non-contingent responses* (corresponding to type III, considered a negative measure) mean = 2.8%, sd = 3.5%; and *missing responses* (corresponding to type IV, considered a negative measure) mean = 12%, sd = 7% (Pagmar et al., 2022). The results showed a conversational conduct where children at the age of 5;0 failed to acknowledge half of their interlocutor's turns.

In light of the studies on linguistic and extra-linguistic development presented above, we were inclined to ask whether or not these proportions would increase during the last years of preschool. We could assume a traceable development of increasing positive conduct, like the increase of Coherence and Report subscales in the CCC results in Yliherva et al. (2009). This result would also be in line with findings on communicative development from Wanska and Bedrosian (1985). However, the conduct of acknowledging one's interlocutor's turn could also stay roughly the same between ages 5;0 and 7;2, similar to the Use of Context subscales, also from the CCC results in

Yliherva et al. (2009) or frequencies of different demonstratives in Turkish children (Küntay & Özyürey, 2006). There is no previous descriptive account of the development of acknowledging one's interlocutor's turn. The current study is a first step in providing such an account.

We made additional observations with the same participants as Pagmar et al. (2022) and the same conversational analysis was conducted.

1.2 Conversation and core language

The term *core language* is used here to refer to a child's productive and/or receptive vocabulary and grammar skills. Early measures of vocabulary show a strong correlation with later language skill (Lee, 2011; Rowe et al., 2012; Song et al., 2015). Assessments of vocabulary and grammar are often correlated during development (Fenson et al., 1994; Hoff et al., 2018; Wilson & Bishop, 2022).

Core language has also been connected to several pragmatic abilities in both typical (e.g. Foppolo et al., 2021; for overview see Matthews et al., 2018) and atypical development (Capps et al., 1998; Hale & Tager-Flusberg, 2005). In atypical development, these findings are specifically connected to conversational behaviour, providing additional indications of a strong relationship between core language and conversational conduct.

Knowing the meaning of words is likely to facilitate the child's focus on other/social aspects of a conversation. This is also the interpretation made by Pagmar et al. where receptive vocabulary at 4;0 predicted two conversational measures at 5;0 (Pagmar et al., 2022): it was a positive predictor for child responses that acknowledged the previous turn and a negative predictor for *missing turns* (instances where +2 sec passed without the child offering any response to their interlocutor). The suggested explanation for this relationship is that early word-learning enables an individual to focus on social aspects of linguistic communication to a higher degree. Therefore, it was of interest to examine if receptive vocabulary at 4;0 still predicted conversational conduct beyond the age of 5;0. If it did, it would provide further indications that early receptive vocabulary lays a foundation for later conversational conduct.

Previous results show no relationship between productive morphosyntactic accuracy and conversational outcome measures[2] (Pagmar et al., 2022), which in a sense contrasts with the results for receptive grammar in Wilson and Bishop (2022). The relationship between conversational conduct and morphosyntactic ability must therefore be examined further.

1.2.1 Core language in the current study

The participants in the current study were part of a pre-existing longitudinal dataset, the MINT project. They were the same individuals that were observed at 5;0 in Pagmar et al. (2022).

[2] Pearson's R from Pagmar et al. (2022) for morphosyntactic accuracy vs. appropriate responses = 0.06.

Following the previously found evidence of a relationship between receptive vocabulary at 4;0 and conversational conduct at 5;0, we wanted to extend the longitudinal gap and examine if receptive vocabulary was still a positive predictor of conversational conduct at 7;2. Observations of receptive vocabulary, PPVT, were part of the pre-existing longitudinal dataset, in which all subjects in the current studies were participants. In line with the evidence from the previously found relationship between receptive vocabulary and appropriate conversational responses, we expected that measures of the children's receptive vocabulary at 4;0 would provide a longitudinal positive prediction of appropriate conversational responses at 7;2, and a longitudinal negative prediction of missing conversational responses at 7;2.

Pagmar et al. (2022) found no connection between conversational conduct and productive morphosyntactic accuracy at 5;0. To follow up on this result, we also tested the relation between a contemporary measure of morphosyntactic accuracy and specific conversational responses. Following Pagmar et al., we expected no reliable relationship between morphosyntactic accuracy and conversational responses. On the other hand, children at age 7;0–11;0 in Wilson and Bishop (2022) showed connections between grammatical skill and several pragmatic abilities. Children in the current study were all age 7;2. It is possible that morphosyntactic accuracy and conversational responses will be connected, in line with Wilson and Bishop. Core language abilities (vocabulary and morphosyntactic accuracy) are not necessarily dependent on socio-communicative skills (see e.g. Kissine, 2021). A strong predictive relationship between core language and our conversational measures could suggest that conversational conduct is less reliant upon social knowledge and taking another person's perspective and more affected by formal language skill.

1.3 Conversation and executive functions

In this section, the general relationship between language use and executive functions (EF) is described briefly, followed by questions concerning conversation and executive functions specifically.

It has previously been suggested that conversational conduct is aided by interlocutors relying on a set of EF that govern parts of this conduct, such as sufficient attention-span, cognitive flexibility, working memory, shifting perspective, updating, and inhibition (Brocki & Bohlin, 2004). During development these abilities are often treated as one unitary cognitive function (Fuhs & Day, 2011; Wiebe, et al., 2008).

Studies on the developmental relationship between EF and pragmatic ability in general, and conversational ability in particular, are sparse (Bendtz, et al., 2022; Blain-Brière et al., 2014; for overview see Matthews et al., 2018). As previously stated, Blain-Brière et al. (2014) found that the variable "talkativeness" (the number of utterances during a semi-structured conversation) showed a negative relationship to measurements of inhibition, while "responsiveness" (the

number of utterances that acknowledged a conversational partner's turn) showed a positive relationship to measurements of working memory. When examining children's pragmatic ability more generally (ages 4;0–5;0), Nilsen and Graham (2009) found no connection between their cognitive flexibility and pragmatic production or comprehension. These findings, together with the ones from Blain-Brière et al. suggest that executive functions may be connected to conversational contingency but not to pragmatic competence in general, which agrees with the second possibility (see Section 1, outcome 2), that conversational conduct is not mainly dependent on pragmatic inferences, but rather on keeping track of information. In Pagmar et al. (2022), an early measure (obtained at 2;9) of short-term memory showed no predictive relationship for conversational outcome measures at 5;0. Due to the young age of the participants, this result was interpreted as representing knowledge of numbers rather than short-term memory.

1.3.1 Executive functions in the current study

In the current study, cognitive flexibility was assessed through the Wisconsin Card Sorting Task (WCST) (Lehto, 1996), and working memory through Backward Digit Span (BDS).

The WCST provides two groups of measures: a group of global measures and a group of measures on perseverative performance. The global measures have been shown to highly correlate with working memory measures (Lehto, 1996; Stratta et al, 1997). Perseverative errors have previously been used for testing cognitive flexibility in general (Van Eylen et al., 2011) although they have also been shown to be highly correlated to measures of visual working memory in children (Huizinga, et al., 2006). Previous research has shown that cognitive flexibility correlates with aspects of productive language use (i.e. quality of descriptions in Bacso and Nilsen, 2017), but connections to conversational conduct have not previously been tested. If WCST shows connection to the conversational measures, it would strengthen the idea of cognitive flexibility as part of productive language use.

One ability that is needed in a conversation is being able to keep track of conversational topics. Conversationalists need to link new turns to previous ones. A positive correlation between conversational conduct and working memory tasks would thus be expected. Following Blain-Briere et al. (2014) we could also expect working memory to correlate negatively with the lack of the same conduct. If Nilsen and Graham's model holds, in which general pragmatic competence did not correlate with cognitive flexibility (2009), then we would expect no strong relationship between perseverative errors and conversational responses. On the other hand, if conversation requires keeping track of different speaker goals (inhibiting some, switching to others), then cognitive flexibility may have a larger role to play. Predictive relationships between measures of EF and conversational conduct would provide an indication that conversational conduct is aided by swift updating of one's beliefs in accordance with new information and strong online representations of conversational content.

1.4 Conversation and specific pragmatic abilities

Pragmatic ability can be described as a multifaceted online capacity to govern the inference of meaning (intentional content) in both receptive and productive communicative signals. Several more or less specific abilities are categorised as part of pragmatic ability, e.g. indirect language use and conversational implicatures, referential production and resolution, altering signals in relation to different interlocutors, narrative ability, etc. Most of these abilities share a common feature: they are used to prompt or obtain intentional content. With this said, there is no or little empirical data suggesting that pragmatic ability must be understood as one developmental ability. On the contrary, Wilson and Bishop (2022) assessed the pragmatic comprehension abilities of 400 children, ages 7;0–13;0, and the result showed quite minor correlations between tests of specific pragmatic abilities[3]. They concluded that task-specific skills played an important role in the children's performances and suggested that pragmatics at large might be best understood as a set of different skills, rather than a specific domain, and that different pragmatic skills can stem from different cognitive underpinnings.

Specific pragmatic abilities of children, in the age span of interest for the current study, have been widely studied during the last few decades. One example is children's comprehension of scalar inferences (Foppolo, et al., 2021; Katsos, 2014; Noveck, 2001; Pagmar et al., in prep; Papafragou & Tantalou, 2004). When turning to the literature, it is apparent that studies on connections between pragmatic skill and other cognitive abilities have been much less studied and then mainly focused on pragmatic comprehension (e.g. Wilson & Bishop, 2022), with less attention directed towards pragmatic production. Therefore, in the current study, we test three abilities of pragmatic production (referential production, audience design, and narrative skill), as well as one of pragmatic comprehension (conversational implicatures).

1.4.1 Pragmatic abilities in the current study

In the current study, we examined the relation between conversational conduct and four aspects of pragmatic comprehension and production. The first aspect was the comprehension of *conversational implicatures* (Grice, 1975), i.e. the conduct of stating something by saying something else, e.g. saying *It's rather cold in here, isn't it?*, when you want someone to close the window. Conversational implicatures were tested through an adapted version of a task constructed by Wilson and Bishop (2019). The second ability was *referential production in a multiple features task*, which measures the conduct of assigning sufficient and necessary information when marking referents on the basis of e.g. temporal, spatial, or general knowledge factors. A similar task has previously been conducted by Dahlgren and Sandberg (2008). The third ability: *audience design* (AD) is the conduct of adjusting the communicative signal with the conversation partner/

[3] Wilson and Bishop (2022) tested a novel online assessment battery with four separate tests of different pragmatic abilities: Implicature Comprehension Test, Children's Test of Local Textual Inference, Pragmatic Violations, and Social Overtures.

audience in mind. AD is also a productive pragmatic ability and was measured with a novel design (Arvidsson et al., 2022; Pagmar et al., 2021). Lastly, the fourth ability considered was *narrative skill*, the ability to follow conventional form, e.g. chronology, when presenting a sequence of utterances for the purpose of representing connected actions and/or events. Narrative skill was tested through an adapted version of the MAIN test (Gagarina et al., 2019). Narrative production can be thought of as less of an interactional component when compared to the other abilities. However, it is a productive pragmatic ability that relies on inferential processes (e.g. Bohnacker & Gagarina, 2020; Filiatrault-Veilleux et al., 2016).

All pragmatic measures included in the current study entail inferential processes, although of different kinds. For a conversationalist, a conversational implicature entails making inferences about what someone else wants to communicate, while audience design entails making inferences about what someone else knows. If conversational conduct mainly relies upon making pragmatic inferences, in line with our first suggestion of outcomes (O1), then we expect to see correlations between these measures and the conversational measures.

Three productive tasks were chosen, since production has been less examined than comprehension when it comes to pragmatic skills (and language skills in general). The inclusion of the Implicature Comprehension Task, also allows for investigation of whether productive and comprehensive tasks converge. Without making a claim concerning the factors that govern conversational conduct, it is true that the act of providing contributions to a conversation is a productive task. Therefore, it is of interest to see how it relates to other productive abilities.

Among the reasons for why we should entertain the idea of a discrimination between pragmatic production and comprehension, there is a asymmetry account: Mognon et al. (2021) concludes that the comprehension of *some* requires the hearer to consider the speaker's perspective, but never the other way around. The speaker is not considering the hearer and this could in turn explain results linking comprehension, but not production, of scalar terms to the development of theory of mind (2021: 12).

All these aspects of language use have the potential to affect the conversational conduct of the developing child, but to what extent they are connected is unknown. If acknowledging previous turns relies more on drawing inferences about the intentions of others, then we would expect agreement between comprehension of conversational implicatures and positive conversational conduct. As an additional exploratory investigation, beyond connection to conversational conduct, we also analysed how the specific pragmatic measures related to one another, as well as to core language and EF.

1.5 Summary

Throughout Section 1, we have outlined a series of questions connected to the development of conversational conduct, in terms of acknowledging one's interlocutor's previous turn, and how it is connected to other developmental abilities. We have presented the sometimes sparse literature

that directly or indirectly relates conversational conduct to core language ability, executive functions, and specific pragmatic abilities. To summarise our expected findings for all these abilities, we list them below in the same order that is employed throughout the paper:

Conversational conduct

We examined conversational conduct longitudinally, comparing proportions of individual childrens different types of conversational responses between 5;0 and 7;2. We expect the proportion of responses that acknowledge one's interlocutor's previous turn to increase at group level, based on previous findings (e.g. Yliherva et 2009). No previous study on children, to the best of the authors' knowledge, has examined this development longitudinally with conversational data.

Conversational conduct and core language

We investigated the relationship between core language (receptive vocabulary and morphosyntactic accuracy) and conversational conduct. We expected receptive vocabulary observed at 4;0 to predict conversational conduct at 7;2., and this would indicate that early receptive vocabulary is a facilitator for later conversational conduct. There were two possible outcomes for concurrent morphosyntactic accuracy: (1) morphosyntactic accuracy would show no reliable connection to conversational conduct, in line with previous findings at 5;0 (Pagmar et al., 2022), or (2) it would correlate positively with conversational conduct, in line with findings of connections between grammatical knowledge and receptive pragmatic abilities for ages 7;0 and above (Wilson & Bishop, 2022). A connection between morphosyntactic accuracy and conversational conduct could be held as an argument for conversational conduct relying more heavily on development of formal knowledge as the children enter school.

Conversational conduct and executive functions

We investigated the relationship between aspects of executive functions and conversational conduct. We did not propose a strong directive hypothesis between the WCST and conversational responses, following e.g. Nilsen and Graham (2009). However, no previous study, to the best of the authors' knowledge, has investigated children's executive functions and conversational conduct specifically.

We expected a positive relation between working memory and appropriate responses, in line with Blain-Brière et al. (2014). A positive connection between conversational conduct and working memory would strengthen the hypothesis that conversational conduct relies more on keeping track of communicative content, rather than just being good at making inferences.

Conversational conduct and specific pragmatic abilities

We investigated the relationship between conversational conduct and four specific pragmatic abilities (comprehension of conversational implicatures, referential production, audience

design in production, and narrative skill in production). The outcomes will be treated as a first step in answering the question concerning what is more important during the development of conversational conduct: inferential processes and social aspects of the interaction, abilities that allow for structural aspects of interaction to be facilitated, or both.

2. Method

Parts of this section have previously been published in a pre-registration on Open Science Framework.

2.1 Participants

All children were participants in the longitudinal project MINT (Gerholm, 2018). The initial conversational measures were obtained for a sample of 40 children at 5;0 (recorded within two weeks of their birthday). The later conversational measures and all other measures of pragmatic skill, executive function, and grammar were obtained from a sample of 50 Swedish speaking children, all at the age of 7;2 (recorded within two weeks of the stated age). Receptive vocabulary was previously obtained at the age of 4;0, for all of the 50 subjects that participated at 7;2. Thirty-four of the subjects participated at both the 5;0 session and the 7;2 session.

2.2 Experimental procedure

All measurements were collected in a laboratory setting at the Department of Linguistics, Stockholm University. All test sessions were video and audio recorded. Two experimenters took turns carrying out the different tests. The tests were carried out on a laptop with the screen facing the child. An experimenter sat beside or opposite from the child throughout each test session. For one test (Backward digit span), the experimenter had an additional laptop to annotate the results in real-time. After all tests had been completed, one of the test leaders (the first author) and the child had a >10 min conversation. Tests were implemented through PsychoPy, PsyToolKit, or delivered manually. Test scores were obtained through the utilised software or through manual annotation of the recorded data in ELAN (2022).

2.3 Tests and measurements

Tests and measurements are outlined below in the following order: conversational measures, tests of core language ability, tests of executive functions, and lastly, tests of specific pragmatic abilities.

2.3.1 Conversational contingency

The experimenter sat across from each participating child for a >10 min conversation. The same experimenter (the first author) carried out all conversations. The conversations revolved around

starting first grade. To make the conversations as uniform as possible while still allowing for free interaction, the test leader produced 12 predetermined utterances (Appendix 2), spread out over the 10 minutes of interaction. All children had met and interacted with the experimenter on several previous occasions. The conversational data was coded in accordance with the coding scheme for Conversational Contingency (Abbot-Smith et al., 2021). Every turn that the child took in response to the researcher during the conversation, both following the predetermined utterances and during the free interaction, was categorised as follows:

contingent,
defined as an appropriate, informative and on-topic response to the experimenter's prior utterance or question.

non-contingent,
defined as utterances that do not maintain the topic of the previous utterance or question.

minimal response,
defined as utterances with little semantic weight, such as "Yeah" or "Wow". One-word utterances are normally coded as minimal.

other,
defined as responses on the part of the child or the experimenter that do not fit into any of the other categories.

missing response,
defined as the occurrence of the child not providing any vocal or gestural response for >2 sec. following the experimenter's turn, and the experimenter once again taking a turn.

An additional measurement was obtained by adding together responses labelled *contingent* and *minimal response*. This additional measure contained all responses in which the child directly acknowledged the experimenter's previous turn. We call this measurement *appropriate responses*.

The outcome measurements consisted of the proportion of responses, calculated by taking the number of the categorised responses and dividing each by the total number of responses during the conversation. Each of the three outcome measures (appropriate responses, non-contingent responses, and missing responses) will be analysed separately.

2.3.2 Tests of core language

2.3.2.1 Receptive vocabulary

The Peabody Picture Vocabulary Test, PPVT-4 (Dunn & Dunn, 2007) is a task that is widely used for assessing receptive vocabulary. The test was conducted at 4;0. The test was adapted for Swedish participants (Ahlström & Ljungman, 2011), and since the test has not been standardised on a Swedish sample, raw scores were used in the analyses.

2.3.2.2 Morphosyntactic accuracy

Morphosyntactic accuracy was gathered from the narrative production data (2.3.4.4). It was measured through an adapted version of a core language skill scoring scheme (Gerholm et al., 2019; Tonér & Gerholm, 2021), which provides an accuracy score, calculated as % well-formed clauses. This measure has previously been used to represent grammatical skill (Pagmar et al., 2022).

2.3.3 Tests of executive function

2.3.3.1 Wisconsin Card Sorting Task

The test was carried out on a laptop. Four images of objects were displayed on the screen. They differed in shape, number and colour. In the lower left corner, an image appeared that had either the same shape, colour or number of objects as any of the four images the child had already seen. The child's task was to match the lower left image with one of the four images, based on a rule of combination unknown to the child. This was done through trial-and-error and if the match was correct, a specific sound confirmed the match. If the match was incorrect, there was a sound telling the child that the match was wrong. The child was instructed to continue with one strategy, once they had found the right match. After a number of trials, the rule of combination changed, and once again there was a sound telling the child that the match was wrong. They then had to change their strategy and sort the images according to one of the other two possibilities (shape, colour, number). The test provided important information about the child's ability to keep in mind which quality (colour, shape, number) was last used when it was time to change strategy, as well as the child's ability to change strategy when required.

We used the raw score of perseverative errors (the number of occasions when the child continues with the same response strategy following a rule switch during the test) as the measure for the analyses.

2.3.3.2 Backward digit span

Backward digit span is used as a measure of working memory (e.g. Nilsen & Graham, 2009). The test was conducted by an experimenter, sitting across from the participant. The experimenter read a series of non-chronological numbers, one at a time, and the participant was asked to repeat the numbers in reverse order. The length of the series of numbers was increased by one digit every second series. If the participant made two errors in a row, the test was terminated. The number of correct series was used as the measure for the analyses.

2.3.4 Tests of Pragmatic Abilities

2.3.4.1 Multiple features task

The task stems from "the referential communication task" (Dahlgren & Sandberg, 2008). The experimenter presented sets of four pictures to the child. In each set, the pictures shared several

different characteristics. The child was asked to describe specific images for the purpose of pointing them out to a third party (a fictional person). There were four potential characteristics that were altered between the pictures. Each child was presented with a total of eight sets of pictures. To make the task age adequate, four pictures per set instead of the original 16 per set were used, following a pre-test (n = 5).

Verbal answers for the *Multiple features task* were coded in accordance with Dahlgren and Sandberg's efficiency score (2008: 339): "total number of mentioned relevant features minus total number of irrelevant features, divided by maximum possible number of relevant features". We refer to this measurement as the *Multiple features task*.

2.3.4.2 Audience design

The audience design task was designed to investigate/assess the child's ability to take an interlocutor's frame of reference (or "world knowledge") into account. The child was presented with sets of four pictures of figures, animals, and people. The child was asked to refer to specific pictures, with characters who were well known to the child (controlled by pre- and post-test surveys), in order to point them out to a third party. The third party, a fictional elderly lady, who was described to the child as "very old", did not share the child's frames of reference regarding cultural phenomena and proper names of figures. The child was told to record messages to the elderly lady, so that she could identify specific pictures when she was presented with the same set of pictures at a later occasion. For example, a child was asked to say something so that an elderly lady could identify an image of the fictional character Steve from the videogame *Minecraft*. The child needed to assess the frame of reference of the elderly lady to provide sufficient descriptions. To describe characteristics that could identify the image regardless of world knowledge would be compatible with proper use of audience design. To only use the name *Steve,* of *Minecraft,* would not be compatible with proper use of audience design, since the child is not justified in assuming that the elderly lady knows about *Minecraft*. Each child was presented with a total of 12 sets.

A pre-test survey was used to corroborate that the stimuli used corresponded to a typical seven-year-old's frame of reference (n = 24). A list, compiled from the survey answers, was shown to the parent of the participating child at the beginning of the test session. The parents were asked to mark the referents that they knew that their child was familiar with. The experimenter then assembled a participant-specific test, based on the parents' answers. After the test, the child was once again shown all pictures of the relevant referents and was asked to name each one. Trials where the child did not know the name of the referent (30 out of 803 trials) were excluded from further analysis.

We divided the group by asking: throughout these twelve trials, did the subject at any time exemplify a behavioural response that was in line with appropriate audience design, "yes" or "no". The reason for making this division was to separate the sample into easily interpretable

groups. What was obtained was a dummy variable: did show signs of AD = 1, did not show signs of AD = 0. The measure divided the sample in half (51% = 1, 49% = 0).

2.3.4.3 Implicature comprehension test

The test was based on the implicature comprehension test (Wilson & Bishop, 2019). Participants watched short films containing question-answer pairs. In the test condition, the answers in the film generated an implicature, meaning that the implied meaning of the answer did not correspond to its literal meaning and had to be accessed through inference. The child was then asked a direct question to determine the speaker's intention. Example: The girl in the video says, "*Should I increase the heat in this room?*". The boy in the video answers, "*I just took off my jumper*". The child is asked the following question: "*Do you think the boy wants the girl to increase the heat in the room?*". If the child comprehends the conversational implicature, they should respond "no", which would be registered as the correct answer. The child was presented with three options when answering: "yes", "no", and "I don't know". The test measures individual differences in the ability to understand conversational implicatures.

A total of 20 videos were shown to each participant (18 experimental trials in total; 6 yes, 12 no trials). All videos were translated from English to Swedish. To exemplify and establish "I don't know" as a valid answer from the child's point of view, the participant was shown two initial videos that contained uninformative answers (i.e. a boy says, "*Do you think your brother will come?",* and a girl replies: "*I can't remember what he said"*). All trials consisted of a subset of the films created by Wilson and Bishop (2019). The subset was chosen based on ease of translation into Swedish, if the implicature was coherent in a Swedish context, as well as age-adequacy. Analyses were carried out with the number of correct responses.

2.3.4.4 Narrative production

The task is part of a larger test battery, Multilingual Assessment of Narratives (Gagarina et al., 2015). Two experimenters took part in the testing. Initially one experimenter and the child sat across from each other in the recording room. The experimenter handed the child a small book, containing six pictures, presented two at a time, and the participant was asked to look through the pictures. The pictures were ordered to convey a specific story structure. The experimenter then asked if the participant could tell this story to a second experimenter. The child was told that the second experimenter had never before heard or seen this specific story. The second experimenter entered the room, while the first experimenter stepped outside, and the participant told the story to the second experimenter. We used two experimenters so the child could address a listener that had no previous experience with the story material (at least to the best of the child's knowledge). In doing so, the child needed to present referents and events in an appropriate way for the story to make sense.

The narrative production task is coded in accordance with the Multilingual Assessment Instrument for Narratives (MAIN) coding scheme for free narrative production (Gagarina, et al., 2019). The narrative data is assigned a score between 0 and 17, where a higher score represents a more detailed and conventionally adept narrative structure.

2.4 Statistical analyses

In the following subsections, we present the specific statistical analyses used to answer the research questions and the order in which they were conducted. All analyses were conducted with the programming language R (version 4.1.2). The full R-script, including all analyses, can be found at OSF.

2.4.1 Initial data quality evaluation

Each measure was evaluated for normality through a Shapiro-Wilk's test. If the test showed a significant result, the measure was rank transformed using a rank-based inverse normal transformation using the `RankNorm` function in the RNOmni package (McCaw, 2022). The measures that showed significant Shapiro-Wilk's test results included: *non-contingent responses, missing responses, multiple features task, MAIN, backward digit span, morphosyntactic accuracy*. Outliers were defined as observations beyond 1.5 interquartile range below the first quartile or above the third quartile. No outliers were found in the data.

2.4.2 Longitudinal conversational conduct

To answer the question concerning the development of conversational contingency in a longitudinal sample, we compared the outcome measures from the conversational data for all subjects that participated in both the 5;0 and the 7;2 session ($n = 34$). We conducted a paired t-test to compare percentages of appropriate responses, non-contingent responses, and missing responses in the participants' conversational contributions and plotted individual changes between the two ages.

2.4.3 Conversation skill vs. core language, executive functions, and pragmatic development

The following analyses were conducted by using the `chart.Correlation` function in the PerformanceAnalytics package (Peterson & Carl, 2020).

To answer the question concerning the relationship between conversational conduct at 7;2 and core language ability, we performed a correlation analysis, obtaining Pearson's R for each combination of the three conversational measures and the two measures representing core language skill: PPVT (longitudinal measure from 4;0) and morphosyntactic accuracy (concurrent measure).

To answer the question concerning the relationship between conversational conduct at 7;2 and executive functions, we performed a correlation analysis, obtaining Pearson's R for each

combination of the three conversational measures and the two measures representing EF: WCST and backward digit span.

To answer the question concerning the relationship between conversational conduct at 7;2 and other pragmatic abilities, we performed a correlation analysis, obtaining Pearson's R for each combination of the three conversational measures and the four measures representing different pragmatic skills: implicature comprehension test, audience design, multiple features task, MAIN. For the binary measure audience design, we present the biserial correlation coefficient.

The results were summarised in three correlation matrices, with correlation coefficients, distribution histograms, and scatter plots. For an overview of the measures included, see **Figure 1**.

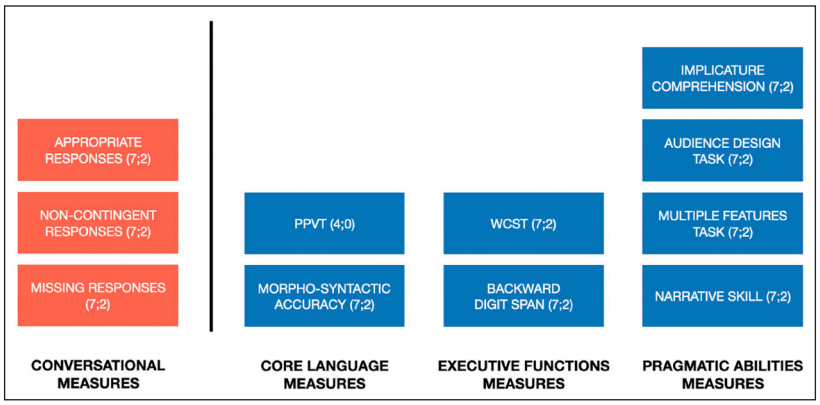

Figure 1: The conversational measures (RED) and the measures of core language, executive functions, and specific pragmatic abilities (BLUE). All measures were obtained at age 7;2, except for the measure of receptive vocabulary (PPVT) which was obtained at age 4;0.

2.4.4 Bayes factor

We conducted additional analyses to obtain Bayes factor values for non-significant correlations. This was done to determine whether these potential relationships showed evidence for the null hypotheses. If the different abilities had been connected, positive correlations between different measures of pragmatic ability and other core language and cognitive measures were expected throughout. Likely alternative outcomes are no correlations or weak negative correlations. In the case of a weak negative correlation, our Bayesian testing could establish evidence that two measures are "not positively correlated" (rather than "potentially negatively correlated"). The analysis was carried out using the `correlationBF` function in the BayesFactor package (Richard & Rouder, 2021), calculated with wide prior scales (1/√3).

2.4.5 Generalised linear models

We decided to conduct a complementary statistical analysis using a predictive framework. We fitted a generalised linear model (GLM) for each of the three conversational measures included

in the current study. The analyses were conducted by using the `glm` function in the `stats` library in R (R Core Team, 2022). The conversational measures (appropriate, non-contingent, and missing responses) were each held as the dependent variable in three separate analyses. The measures were gathered from the same conversational data and are therefore not independent of one another. However, each measure represents a conceptually different aspect of conversational conduct, which justifies an individual examination of each of them (for further discussion see Pagmar et al., 2022). The measures of core language (PPVT and morphosyntactic accuracy), executive functions (the WCST and Backward digit span), and other pragmatic abilities (ICT, Audience design, Multiple features task, and Narrative skill) were held as independent variables.

2.4.6 Evaluation of the data for factor analyses

We conducted Bartlett's test of sphericity and a Kaiser-Meyer-Olkin (KMO) test for sampling adequacy, to control for sufficient redundancies in the data that would allow for factors to be defined. We decided that factor analysis would only be considered as an informative method if the Bartlett's test of sphericity was significant and the KMO test showed a $>.6$ result.[4]

3. Result

Descriptive statistics of raw scores for the three conversational measures, as well as measures of core language skill, executive functions, and specific pragmatic abilities are presented in **Table 1**.

Table 1: Descriptive statistics of raw scores for both dependent and independent variables, including the mean, standard deviation, and minimum and maximum value.

	Measure	Mean (SD)	Min.	Max.
CONVERSATIONAL MEASURES	Appropriate responses	75.8 (11.2)	39	107
	Non-contingent responses	1.4 (1.6)	0	7
	Missing responses	9 (6.1)	0	28
CORE LANGUAGES	PPVT (at 4;0)	33.4 (7.1)	19	51
	Morphosyntactic accuarcy	3.7 (0.3)	2.59	4
EXECUTIVE FUNCTIONS	WCST	15.3 (3.5)	8	23
	Backward digit span	3.2 (0.9)	2	5
PRAGMATIC ABILITIES	Implicature comprehension test	11.7 (2.3)	7	18
	Multiple features task	6.2 (1.5)	1.5	8
	Narrative skill	5.9 (2)	2	12

[4] Neither Bartlette's test of sphericity nor the KMO test met the criteria for further analysis. The test results, as well as the results of the analyses can be viewed in Appendix 1.

The binary measure from the *audience design* test split the sample in half, with 49% of subjects consistently naming all referents (an answer not compatible with proper audience design and therefore assigned the value 0), and 51% of subjects offering at least one description of a referent that they knew the name of (and therefore assigned the value 1).

3.1 Development of conversational responses

The analysis addresses the longitudinal examination of conversational responses (presented in 1.1.1). The analyses of the conversational data show an increase of the positive conversational measure, and a decrease of the negative conversational measures, between 5;0 and 7;2. Thirty-four of the participating subjects were observed at both ages, and the longitudinal change of conversational conduct for these subjects is displayed in **Figure 2**. A paired sample t-test showed a reliable increase of appropriate responses, $t(33) = -5.8$, $p < .001$, 95% CI [–.22,–.11] (from 44% to 62%), as well as a reliable decrease in both non-contingent responses, $t(33) = 3.02$, $p = .004$, 95% CI [.006,.032] (from 3% to 1.2%), and missing responses, $t(33) = 2.79$, $p = .008$, 95% CI [.011,.075] (from 12% to 8%).

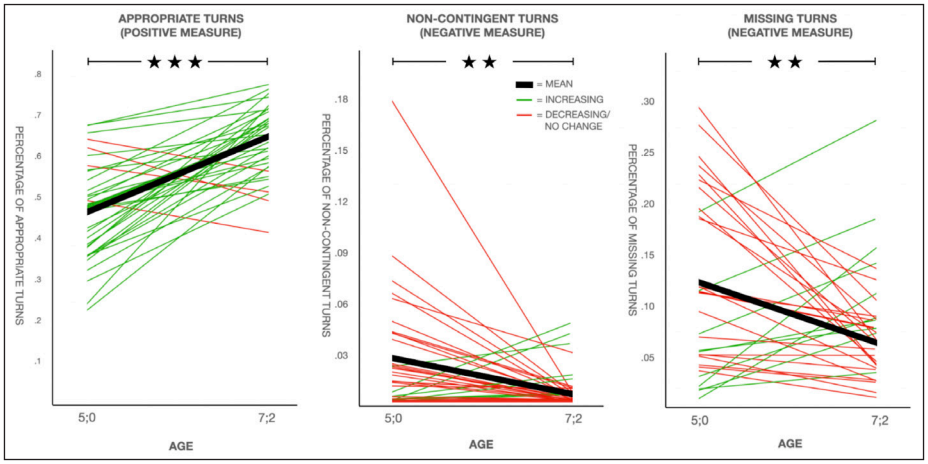

Figure 2: Longitudinal differences in the conversational measures for children at 5;0 and 7;2. Each line represents one individual and the change in percentage of the specific outcome measure between observations. The percentage of appropriate responses (responses in which the child directly acknowledges their partner's turn) shows a reliable increase between the two occasions. There is also a reliable decrease in the percentage of non-contingent responses, as well as for missing responses.

3.2 Correlation analyses

In the following section, we present the correlation analyses between the conversational measures and measures of core language skill (3.2.1), measures of executive functions (3.2.2) , and measures of specific pragmatic abilities (3.2.3).

There is a reliable negative relationship between the number of appropriate responses and missing responses, t(41) = 3.1, p = .003, 95% CI [–.68,–.21]. No significant correlations were found between any of the core language measures and the measures of EF or specific pragmatic abilities. Correlation coefficients between all measures included in the study are included in Appendix 3.

3.2.1 Conversational measures and core language skill

The analysis addresses the relation between core language ability and conversational measures (see 1.2). There were no reliable relationships between conversational measures and the core language measures (PPVT and grammar, see **Figure 2**). Correlation coefficients for all conversational measures and the measures of core language skill are presented in **Figure 3**.

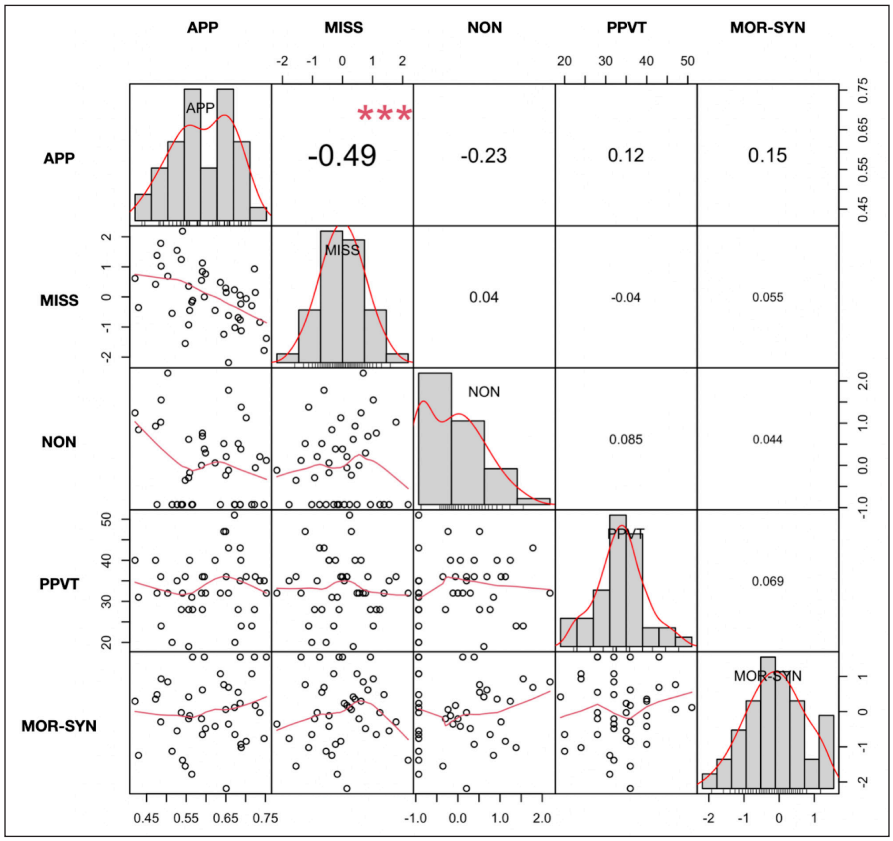

Figure 3: A correlation matrix, with measure histograms and scatter plots, including the three measures representing conversational conduct, appropriate responses (APP), missing responses (MISS), non-contingent responses (NON), and measures representing core language skill, i.e. a longitudinal measure of receptive vocabulary (PPVT), and a measure of morphosyntactic accuracy (MOR-SYN).

The two measures representing core language skill, PPVT and morphosyntactic accuracy, do not show a reliable positive relationship, t(41) = .44, p = .66, 95% CI [–.23,.36].

3.2.2 Conversational measures and executive functions

The analysis addresses the relation between executive functions and the conversational measures (see 1.3). The conversational measure missing responses shows a negative relationship with the WCST, t(41) = 2.09, p = .04, 95% CI [–.55,–.01]. Non-contingent responses showed a negative relationship with the measure backward digit span, t(41) = 2.9, p = .004, 95% CI [–.64,–.13]. No other pairs of conversational and executive function measures showed any significant relationship (see **Figure 4**).

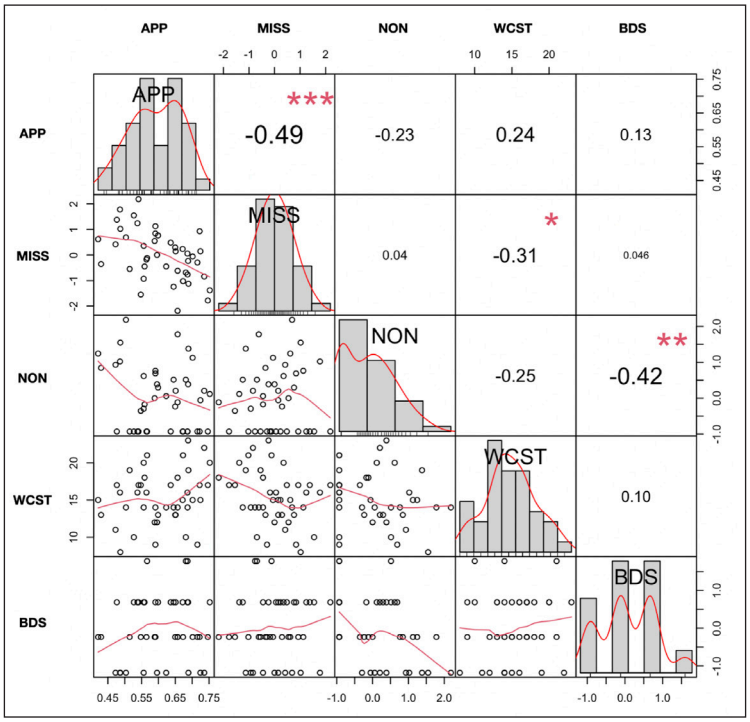

Figure 4: A correlation matrix, with measure histograms and scatter plots, including the three measures representing conversational conduct, appropriate responses (APP), missing responses (MISS), non-contingent responses (NON), and measures representing executive functions, i.e. Wisconsin Card Sorting Task (WCST), and Backward Digit Span (BDS).

The two measures representing executive functions, WCST and Backward digit span, do not show a reliable positive relationship, t(41) = 0.67, p = .5, 95% CI [–.2, .39]. No significant correlations were found between any of the measures of EF and the core language measures or specific pragmatic abilities (Appendix 3).

3.2.3 Conversational measures and other pragmatic abilities

The analysis addresses the relation between pragmatic abilities and the conversational measures (see 1.4). Non-contingent responses show a negative relationship with the comprehension of conversational implicatures (ICT), t(41) = 2.48, p = .016, 95% CI [–.59,–07]. The data does not

show any significant relationship between the other measures of conversational conduct (Appropriate responses and Missing responses), and the measures representing other pragmatic abilities, nor any reliable relationship between the different measures of pragmatic abilities (see **Figure 5**).

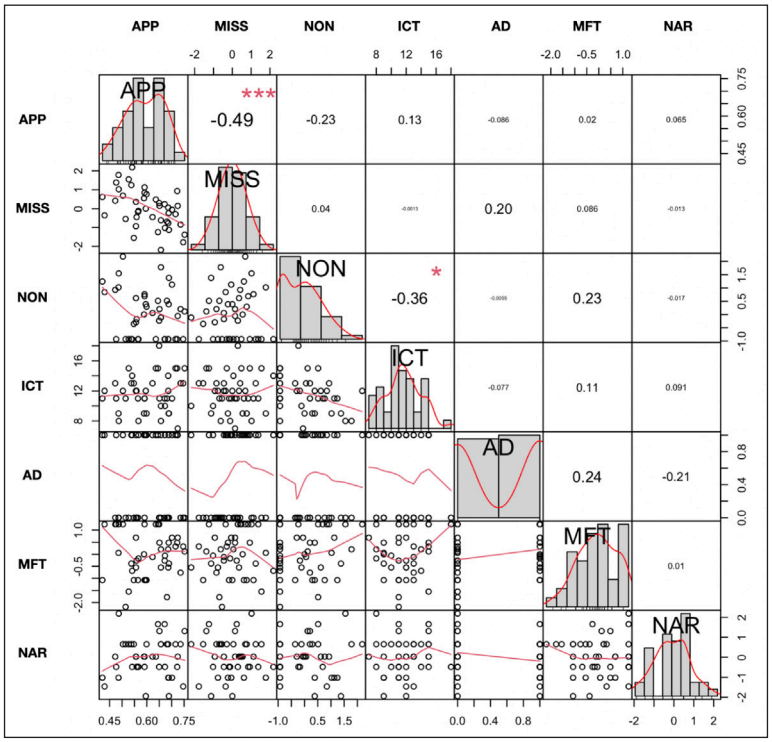

Figure 5: A correlation matrix, with measure histograms and scatter plots, including the three measures representing conversational conduct, Appropriate responses (APP), Missing responses (MISS), Non-contingent responses (NON), and measures representing other pragmatic abilities, i.e. implicature comprehension task (ICT), Audience design (AD), Multiple features task (MFT), and narrative ability (NAR). For AD the biserial correlation coefficient is presented.

3.3 Bayes Factor for non-significant correlations

In **Table 2**, the reader will find the Bayes factor values (BF) for non-significant correlations that provided moderate evidence for the null hypothesis (H0) on no connection between variables. In terms of conventional interpretation of BF, 13 correlations show moderate evidence for H0, and eight correlations show anecdotal evidence for H0.

3.4 Results of the generalised linear models

The GLM analyses showed results very similar to the correlation analyses. There were no significant predictors for appropriate responses, two significant predictors for non-contingent responses (the EF measure BDS representing working memory and the measure ICT representing

Table 2: Bayes factor results for non-significant correlations with moderate evidence for H0, calculated with wide prior scales (1/√3).

Correlation variables	Bayes factor	Evidence for H0
CORE LANGUAGE SKILL		
Non-contingent responses vs. PPVT	.29	Moderate
Non-contingent responses vs. morph/syn acc.	.26	Moderate
Missing responses vs. PPVT	.26	Moderate
Missing responses vs. morph/syn acc	.27	Moderate
EXECUTIVE FUNCTIONS		
Missing responses vs. backward digit span	.26	Moderate
PRAGMATIC ABILITIES		
Appropriate responses vs. audience design	.29	Moderate
Appropriate responses vs. multiple features task	.26	Moderate
Appropriate responses vs. narrative ability	.27	Moderate
Non-contingent responses vs. audience design	.25	Moderate
Non-contingent responses vs. narrative ability	.25	Moderate
Missing responses vs. ICT	.25	Moderate
Missing responses vs. multiple features task	.29	Moderate
Missing responses vs. narrative ability	.25	Moderate

comprehension of conversational implicatures), and one significant predictor for missing responses (the EF measure representing cognitive flexibility). The complete models are presented in Appendix 4.

4. Discussion

Through longitudinal observations of semi-naturalistic conversations, we have shown that children's ability to acknowledge their interlocutor improves between 5;0 and 7;2. All three types of conversational responses advanced longitudinally in the expected directions: appropriate responses increased, and both non-contingent and missing responses decreased between the observed ages (from preschool years to school start). The longitudinal slopes show the individual developmental transitions (**Figure 1**). A few individual slopes show a direction opposite from what was hypothesised. This is expected in conversational data due to its dynamic nature and can possibly be explained by, for instance, the mood of the child on the day of observation. The results also indicate that the time period between our observations is one of important

behavioural adaptation: non-contingent responses showed a skewed distribution at 5;0, with 30% of children showing a more frequent production of non-contingent responses. At 7;2, the distribution is still skewed, but the spread of the measure is much narrower, meaning that some children still produce non-contingent responses, but not nearly to the same extent as during the observations at 5;0. The mean of missing responses is also lowered, which indicates that the children are adapting to a more conventionalised conversational conduct, but that this aspect of the conversational ability is still developing. This result provides an indication that adult levels of acknowledging one's interlocutor's turn (>85%) are not reached by a sudden realisation of a social norm. Rather, children seem to show a gradual adjustment towards a conversational convention.

We also wanted to study the relationships between the children's conversational conduct at school start and concurrent measurements of their core language skill, executive functions, as well as a set of specific productive and receptive pragmatic abilities. Besides presenting the relationships between the included abilities and conversational conduct, we wanted to find indications of what type of abilities agree with the development of conversational conduct. The two measures of core language skill (longitudinal receptive vocabulary and concurrent productive morphosyntactic accuracy) did not show any reliable connection to any of the conversational measures at 7;2, neither in the correlation analysis nor in the GLM (see Appendix 4). Moreover, the Bayes factor analysis provides moderate evidence of no connection between morphosyntactic accuracy and both non-contingent and missing responses. The results concerning productive morphosyntactic accuracy were thus in line with the hypothesised outcome and replicate the result from Pagmar et al. (2022). The result concerning receptive vocabulary contradicts the hypothesis, which stated that the longitudinal measure of receptive vocabulary that previously showed a strong predictive relationship to conversational responses at 5;0 no longer predicts either appropriate or missing responses. There are several potential explanations for this result, presented below (see section 4.1).

The two measures representing executive functions, WCST and backward digit span, both indicate connections to conversational responses, albeit different ones. Non-contingent responses have a negative relationship to backward digit span performance in both the correlation analysis, $r = -.42$, and the GLM, indicating that children who offered more responses without acknowledging their conversational partner's prior turn also struggled with the working memory task. Missing responses show a negative relationship to WCST performance ($r = -.31$), indicating that children who more frequently offered no responses to their conversational partner's prior turn also struggled with the task of cognitive flexibility and shifting. A possible explanation for the connection to missing responses is that cognitive flexibility allows for adjusting to new information during conversation, and that a child who struggles with this adjustment will opt out of the cooperative activity, rather than moving forward in uncertainty. The results for EF

share similarities with the findings from Blain-Brière et al. (2014), who encountered a negative relationship between "talkativeness" and inhibition, as well as a positive relationship between "responsiveness" and working memory, at age 4–5. Both of these results and connections to previous findings are discussed in more detail below (see 4.2).

Only one of the tested pragmatic abilities showed indications of a more reliable connection to one of the conversational measures: the Implicature Comprehension Task was negatively correlated with non-contingent responses ($r = –.36$). This relationship was also found in the GLM analysis. This indicates that children who offered more responses without acknowledging their conversational partner's prior turn also struggled with understanding indirect language use. This result is in line with the expected outcome. None of the other three measures (audience design, multiple features task, and narrative ability) showed any connection, neither towards the conversational responses or among themselves. The results are discussed in greater detail below (see 4.3). A correlation matrix for all measured abilities can be found in Appendix 3.

We initially entertained three possible outcomes: conversational conduct would be mainly explained by (O1) measures of pragmatic ability, (O2) executive functions, e.g. working memory, or (O3) that measures of pragmatic abilities and executive functions would both be informative for the conversational measure (O3). The effects that were found agree with O3 and with a view of conversational conduct as an ability aided by both the ability to make pragmatic inferences and to track conversational content from previous turns. The outcomes also confirm that conversation during development is a dynamic activity that simultaneously can put strain on several different abilities.

4.1 Core language and conversational responses

4.1.1 Productive morphosyntactic accuracy and conversational responses

The measure of productive morphosyntactic accuracy does not show a reliable correlation to any of the conversational measures. The Bayes factor analysis provides moderate evidence for the null hypothesis between morphosyntactic accuracy and both non-contingent responses ($BF = .26$) and missing responses ($BF = .27$).

We provide moderate evidence that productive morphosyntactic accuracy and conversational conduct are unrelated at age 7;2, replicating the similar previous findings for children at 5;0 (Pagmar et al., 2022). When studying the relation between conversational development and potentially related skills, this is one of the more robust findings yet in the literature. The results suggest a developmental segregation between processes of linguistic form and the use of linguistic signals in social interaction. The result from the previous study was not evaluated in terms of evidence for the null hypothesis, but the correlation coefficient between morphosyntactic accuracy and appropriate responses was 0.06 (2022: 153). The suggested explanation in Pagmar et al. was that (i) a certain level of morphosyntactic ability is necessary initially, if a child is to

maintain a back-and-forth conversation. However, (ii) once a certain morphosyntactic acquisition threshold has been reached, morphosyntax no longer accounts for individual differences in child conversational ability. The results of the current study are in line with this suggestion. When integrating the results, we suggest that processes are segregated before the age of 5;0 and remain segregated, at least until 7;2, and potentially for the rest of the lifespan.

4.1.2 Productive morphosyntactic accuracy and other pragmatic abilities

In contrast to previous results (e.g. Wilson & Bishop, 2022), we found no connections between morphosyntactic accuracy and pragmatic abilities. This could be explained by potential differences between the production and comprehension of grammar at the age of 7;2. Production measures might have an advantage when characterising individual differences of uniquely grammatical processes, as a full syntactic analysis may be circumvented in comprehension, but not in production. As Eve Clark puts it: “logically, comprehension must precede production” (Clark, 1993: 246) and a production test of a certain grammatical construction is thus generally a more difficult test than a comprehension test of the same construction, at a given age. Another possible explanation is other differences between the tasks. In this study, the measure consisted of an accuracy score, calculated as % well-formed clauses during the narrative production task. In Wilson and Bishop, the participants were asked to make grammaticality judgements, which can be argued to be a less ecologically valid task. Language users are not asked to make explicit judgements on whether or not their interlocutor’s utterances are grammatically correct during everyday language use. While language users can easily detect syntactical errors, they also accept a large number of incongruent clauses, repairments, etc. Assessing morphosyntactic ability from material consisting of free language production, in which the child is not explicitly conducting a morphosyntactic task, should therefore be considered a more ecological measure.

4.1.3 Receptive vocabulary and conversational responses

The measure of receptive vocabulary at 4;0 shows no correlation with any of the conversational measures at 7;2. We do find moderate evidence in the BF analysis for the null hypothesis of no connection between receptive vocabulary at 4;0 and both non-contingent responses (BF = .29) and missing responses (BF = .26). This is unexpected, since appropriate conversational responses are increasing and we know that receptive vocabulary is also an ability that increases during the age span studied in the current study. The result provides evidence against the suggestion that early receptive vocabulary is a facilitator for conversational conduct throughout development. We suggest two possible explanations for this result: (1) the abilities that allow conversational conduct to develop rely on aspects of the core language skill only during certain phases of the developing process. Conversational ability might be aided by vocabulary, and potentially vice versa, at early stages of the developmental process of interactional language use, but will be less connected as more normative conduct and conventional behaviour is established. The other

suggested explanation is that (2) the conversational abilities that allow for appropriate responses to increase do develop in agreement with receptive vocabulary, but the developmental slope of the vocabulary undergoes sufficient distributional in-group changes so that vocabulary at 4;0 no longer can account for longitudinal prediction. Although core language measures show long-term stability (e.g. Bornstein et al., 2016), there is empirical evidence that might grant this suggestion some merit. A longitudinal study by Song et al. (2015) investigated the vocabulary growth between 4;0 and 11;0. They identified three rather distinct developmental paths of vocabulary growth: (i) children who started with a high vocabulary assessment and had high assessments throughout the study, (ii) children who started with a low vocabulary assessment and had low assessments throughout the study, and (iii) children who started with a low vocabulary assessment and then showed increasingly higher assessments throughout the study and ended up close to the first group. Although there is a steady increase in receptive vocabulary, relative positions between individuals may change within a sample, indicating different developmental paths. The results suggest that early receptive vocabulary does not facilitate later conversational conduct.

4.2 Executive functions and conversational responses

4.2.1 Non-contingent responses and working memory

The measure of working memory (backward digit span) is negatively correlated with non-contingent responses. This indicates that a child who struggled during the working memory task more often offered conversational turns without acknowledging their interlocutor's previous turn. A possible reason for the non-contingent response is that the previous turn is simply not remembered. This explanation is in agreement with the suggestion of conversational conduct as a conduct that is heavily dependent on keeping track of previous turns and how new turns relate to and further the current topic of the conversation. If we assume that interpretation of an utterance will relate to the content of previous utterances, and that there often will be inferential processes that allow for these interpretations to take place, then it would explain why this aspect of conversational conduct relies on working memory. This result is related to results from Blain-Brière et al., studying ages 4;0–5;0. The connection they found was between the variable "responsiveness", i.e. the number of utterances that acknowledged a conversational partner's turn (similar to our measure), and measurements of working memory (also backward digit span). In studies of the relation between developmental conversational conduct and potentially related skills, this is perhaps the most robust finding yet, as it has been independently replicated. It is interesting to note that the relation between non-contingent responses and working memory is fairly specific, e.g. not extending to the WCST EF measure.

Working memory is not a measure of inferential ability, but inferential content ought to be what working memory is dealing with during conversational conduct. However, it is possible that an individual with poor inferential ability could infer just enough to follow a conversational topic and, due to strong working memory, show appropriate behaviour.

4.2.2 Missing responses and cognitive flexibility

The negative correlation between the measure for cognitive flexibility (WCST) and missing responses suggests that a child that struggled with updating the rules and following new strategies during the card sorting task showed more instances of not offering any response towards their conversational partners' statements and questions. The result provides an indication that cognitive flexibility allows for a child to more easily make contributions to the conversation. This result is reminiscent of the negative connection found at age 4–5 by Blain-Brière et al. (2014) between their variable "talkativeness", i.e. the number of utterances during a semi-structured conversation, and measurements of inhibition. This inhibition might translate into a type of cautiousness for less talkative individuals, where the important factor is to act correctly rather than be polite. If the child in an interactional context is uncertain about what is expected of them, this might place a high load on cognitive flexibility. The child may opt to not do anything at all. For a child that has less of an issue with such a high load on cognitive flexibility, we would then naturally observe relative talkativeness (i.e. fewer missing responses in our case). Our results indicate that the same dynamics are present also at age 7;2. However, an unexpected outcome is that the indications of a relationship between missing responses and cognitive flexibility are not found for the second measure of executive function, working memory, but rather the opposite: the Bayes factor analysis provides moderate evidence for the null hypothesis of no connection between missing responses and working memory (BF = .26). This contrast is discussed further under 4.4.

4.3 Pragmatic abilities and conversational responses

4.3.1 Conversational implicatures and conversation

Among the four measures of specific pragmatic abilities that were included in the study, only comprehension of conversational implicatures showed a connection to one of the conversational measures: non-contingent responses. Looking at the relationship between non-contingent responses and ICT in **Figure 4**, the scatter plot seems to suggest that it is possible for a participant with a lower number of non-contingent responses to position themselves with both high or low ICT scores, and vice versa, but that no participant that scored high for either measure, also scored high for the other. In other words, none of the children that more often produced non-contingent responses excelled at inferring speaker meaning from conversational implicatures. The result indicates that both implicature comprehension and working memory are linked to conversational conduct; individuals that stay on topic are better at simultaneously inferring what is being communicated while keeping track of what has been communicated.

A possible explanation for the fact that the other measures of pragmatic abilities showed no reliable effects is that they are too specific: they test abilities that do not necessarily translate into everyday language use. ICT is also the only test that relies solely on pragmatic comprehension, in

that a participating child does not need to provide any language production of their own to carry out the task, which leads us to our next point of discussion: differences between comprehension and production.

4.3.2 Narrative ability and conversational responses

The fourth measure of pragmatic ability, narrative skill, shows no indication of a connection to any of the conversational measures, and there is moderate evidence for the null hypothesis of no connection between narrative ability and all three conversational responses: appropriate responses, BF = .27; non-contingent responses, BF = .25; missing responses, BF = .25. This result seems to suggest that, even if conversational conduct could partly be explained by conventional patterns of language use, the processes that allow conventional structures in narratives are different from processes that allow contingent conversational conduct. A notable difference is that narrative ability can be displayed without interactional components.

4.4 Limitations

One feature of the results that might at first stand out to the reader is that the measures of core language skill, receptive vocabulary at 4;0 and productive morphosyntactic accuracy at 7;2 did not correlate. However, this can potentially be explained by individual trajectories of vocabulary development within the group between 4;0 and 7;2 (see Song et al., 2015). In other words, there would have been a stronger expectation of a correlation between concurrent vocabulary measures and morphosyntactic accuracy at 7;2. Future studies should thus include additional concurrent productive and receptive measures for core language ability.

Similar to the core language measures, our EF measures do not show a reliable connection among themselves, which is similar to findings by Tonér and Gerholm (2021). Both of our measures were concurrent, but a potential explanation is that aspects differ between tasks. Previously found correlations between EF measures in Huizenga et al. (2006) relied solely on receptive working memory tasks (Wisconsin Card Sorting Task and the Tower of London) and there was no need for participants to engage in language production. Although both tasks in our study rely on verbal instructions, one is visual, the other auditory, and the context in which the tasks take place is also different: one is a game in front of a computer (WCST), the other is a task of sending verbal signals back and forth between the test leader and the subject (BDS). In BDS, many of the interactional components of conversation are also accessible, i.e. turn-taking, meeting gaze, etc. This might have affected the outcome.

Finally, a more thorough discussion on the limitations of the current audience design task can be found here in Pagmar, Arvidsson, and Uddén (2021). Most importantly, the audience design task was updated accordingly (Arvidsson et al., 2022), to allow for a more detailed interpretation, in addition to the binary measure used in the current study.

4.5 Summary

In the present study, we show that the age span from 5;0 to 7;2 is one of notable conversational development, with appropriate responses increasing, and non-contingent and missing responses decreasing. The results indicate that there is a gradual development towards adult levels of appropriate conversational conduct. Previous findings of no connection between productive morphosyntactic accuracy and the conversational measures at 5;0 were replicated when both measures were made at 7;2. The previously found connection between receptive vocabulary and conversational conduct was not encountered at 7;2. These results suggest that the development of conversational conduct as a process is not dependent on core language skill towards the end of the preschool stage. We also present extended evidence of a relation between missing responses and cognitive flexibility, as measured by the WCST, on the one hand, and non-contingent responses, and working memory, on the other hand.

Non-contingent responses were also negatively correlated with the comprehension of conversational implicatures, suggesting that children that struggled with inferring intentional content from contextually dependent utterances were more likely to not engage in the cooperative conduct that is a maintained conversation. Together with the result of the measures of executive functions, these findings suggest that conversational conduct is related to both inferring communicative intentions and being able to keep track of the contributions of others and how these contributions relate to previous turns. Other tests for pragmatic skills showed no connection to the conversational measures, nor between themselves, providing further evidence for a divide between different pragmatic abilities.

5. Conclusion

The current study supports a view of a notable developmental window for conversational conduct, just before children turn 7. The results show a relative segregation between conversational conduct and core language as well as several of the tested pragmatic abilities, except the ability to understand conversational implicatures. Conversational conduct also showed interesting relations to executive functions, in terms of working memory and cognitive flexibility. We suggest that during later stages of preschool, children’s ability to appropriately acknowledge their interlocutor is aided not only by being able to infer the intentions of others based on their utterances, but also by simultaneously keeping track of previous utterances and how they relate to the conversational topic.

Appendix 1. Factor analyses and evaluations

We altered the previously determined factor analysis method from confirmatory to exploratory, due to an insufficient number of observations.

To answer the question concerning potential underlying factors that could contribute to relationships between pragmatic, linguistic, and executive function abilities, we conducted three exploratory factor analyses (presented in detail below), using the `fa` function in the R library psych (2.1.9). In these analyses, the general pragmatic ability is represented by two measures: the main conversational measure, Appropriate responses, representing pragmatic production, and the Implicature Comprehension Test, representing pragmatic comprehension. These measurements were chosen due to previous findings where they both have shown reliable signals, in terms of hypothesis confirming results (Pagmar et al., 2022; Wilson & Bishop, 2019; 2022). The data set from each analysis was evaluated with a Bartlett's test of Sphericity and a KMO test.

The first factor analysis concerned pragmatic ability and core language skill, and was conducted using PPVT at 4;0 (receptive vocabulary), morphosyntactic accuracy at 7;2, appropriate responses (pragmatic production) at 7;2, and ICT (pragmatic comprehension) at 7;2. We specified a two-factor mode.

The second factor analysis concerned pragmatic ability and executive functions. The data frame included the measurements: perseverative error from WCST (executive function), backward digit span (executive function), appropriate responses (pragmatic production), ICT (pragmatic comprehension). We specified a two-factor model.

Lastely, the third exploratory factor analysis was conducted to investigate relationships between all measures of specific pragmatic abilities. We included the following measurements: appropriate responses (conversational conduct), multiple features task (referential production), ICT (implicature comprehension), and narrative ability. We specified a two-factor model.

The evaluation analyses for determining sufficient redundancies and measure of sampling adequacy of each two-factor model showed insufficient results (**Table A**). Due to insufficient redundancy and adequacy, we did not conduct further analyses with the purpose of examining differences between different models, and the results of the factor analyses were not a foundation for discussion. The results of the exploratory factor analyses are presented in **Figure A**.

Table A: Result from the evaluations of redundancies in the variables that allow for factors to be defined (Barlett's Test of Sphericity) and measure of sampling adequacy (KMO Test).

EFA	Bartlett's Test of Sphericity	KMO Test (Overall MSA)
Pragmatics vs. Core Language	$\chi2 = 4.02$, df = 6, p = .67	.59
Pragmatics vs. Executive Functions	$\chi2 = 5.16$, df = 6, p = .52	.49
Pragmatics vs. Pragmatics	$\chi2 = 1.57$, df = 6, p = .95	.54

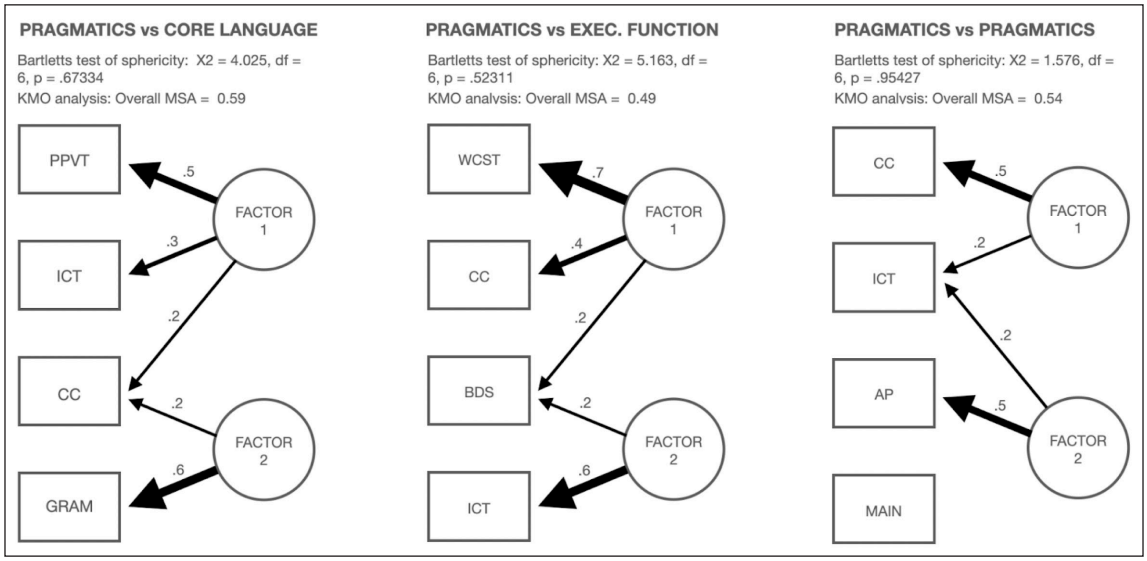

Figure A: The three exploratory factor analyses, presented with the evaluation results from the Bartlett's test of sphericity and the KMO test. The variables are PPVT (receptive vocabulary), ICT (Implicature Comprehension Task), CC (appropriate responses), GRAM (morphosyntactic accuracy), WCST (EF, cognitive flexibility), BDS (EF, working memory), AP (referential production, MTF), and MAIN (narrative skill). MAIN is not connected to either of the two factors specified under the pragmatics vs. pragmatics analysis, since loadings under .2 are not represented with an arrow.

Appendix 2. Predetermined utterances for prompting conversation translated from Swedish to English

1. *You've started first grade*
2. *What are the differences between preschool class and first grade?*
3. *Who's your teacher?*
4. *When I started first grade, I was kind of scared*
5. *I was scared of not getting friends and that the teacher would be mean*
6. *Do you remember these?* [points to stuffed animals]
7. *When I started first grade, we were not allowed to bring stuffed animals to school*
8. *One time at first grade, during lunch, I fell and dropped my food tray, and spilled soup all over myself*
9. *Have you done anything like that?*
10. *Have you ever fallen and injured yourself?*
11. *Is there anything you find hard in school?*
12. NAME-OF-CHILD, *it was so nice talking to you*

Appendix 3. Correlation matrix for all variables included in the study

Here we present Pearson's correlations between all included variables (**Figure B**).

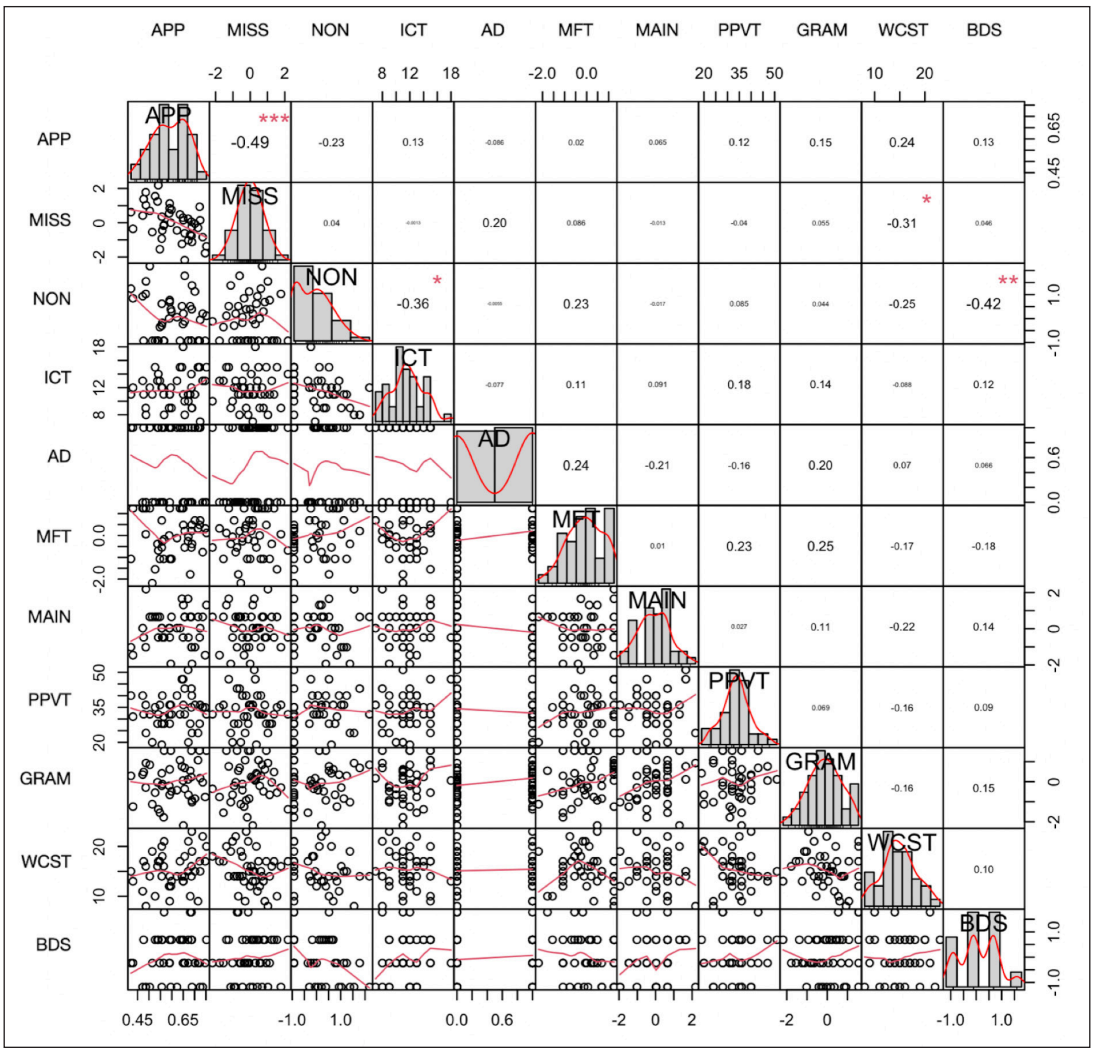

Figure B: Correlation matrix of the three conversational contingency measures (APP, MISS, NON) and pragmatic measures (ICT, AD, MFT, MAIN), core language measures (PPVT, GRAM), and executive function measures (WCST, BDS). Pearson's correlation was calculated for all measures, except for the measure AD, where we calculated the biserial correlation coefficient.

Appendix 4. Generalised linear models

Here we present the Generalised Linear Models for the three measures of conversational conduct. The GLM of appropriate responses showed no reliable effects for any of the predictors (**Table B**).

Table B: The results of the Generalised Linear Model for *Appropriate responses*, in terms of predictor estimates, 95% confidence interval, standard error, t-value, and p-value.

	Estimate	2.5%	97.5%	Std. Er	*t*	*p*
(Intercept)	0.413	0.165	0.662	0.126	3.266	0.002**
PPVT	0.001	–0.002	0.005	0.002	0.646	0.522
MOR-SYN	0.016	–0.014	0.048	0.016	1.029	0.310
WCST	0.007	–0.000	0.016	0.004	1.845	0.073.
BDS	0.005	–0.029	0.040	0.017	0.314	0.755
ICT	0.003	–0.008	0.015	0.006	0.520	0.606
AD	–0.020	–0.082	0.041	0.031	–0.651	0.519
MFT	0.003	–0.031	0.037	0.017	0.178	0.859
MAIN	0.006	–0.024	0.038	0.016	0.417	0.679

<.10 = . ; <.05 = * ; <.01 = ** ; <.001 = ***.

The GLM of non-contingent responses showed reliable effects for two of the included predictors (**Table C**). The measure of working memory, backward digit span (BDS), showed a negative relationship with the proportion of non-contingent responses. The measure for comprehending conversational implicatures also showed a negative relationship with the proportion of non-contingent turns.

Table C: The results of the Generalised Linear Model for n-contingent responses, in terms of predictor estimates, 95% confidence interval, standard error, t-value, and p-value.

	Estimate	2.5%	97.5%	Std.Er	*t*	*p*
(Intercept)	1.982	–0.129	4.093	1.077	1.840	0.074.
PPVT	0.014	–0.021	0.050	0.018	0.799	0.429
MOR–SYN	0.082	–0.186	0.350	0.136	0.600	0.552
WCST	–0.045	–0.116	0.024	0.035	–1.266	0.214
BDS	–0.364	–0.661	–0.066	0.151	–2.398	0.022*
ICT	–0.145	–0.248	–0.041	0.052	–2.750	0.009**
AD	–0.051	–0.575	0.473	0.267	–0.191	0.849
MFT	0.126	–0.166	0.420	0.149	0.846	0.403
MAIN	0.006	–0.260	0.273	0.136	0.045	0.964

<.10 = . ; <.05 = * ; <.01 = ** ; <.001 = ***.

The GLM of missing responses showed a reliable effect for one of the included predictors (**Table D**). The measure of executive function, WCST, showed a negative relationship with the proportion of missing turns.

Table D: The results of the Generalised Linear Model for *Missing responses*, in terms of predictor estimates, 95% confidence interval, standard error, t-value, and p-value.

	Estimate	2.5%	97.5%	Std.Er	*t*	*p*
(**Intercept**)	1.698	–0.994	4.392	1.374	1.236	0.224
PPVT	–0.010	–0.055	0.035	0.023	–0.442	0.661
MOR–SYN	–0.051	–0.393	0.290	0.174	–0.295	0.769
WCST	–0.099	–0.189	–0.009	0.045	–2.173	0.036*
BDS	0.110	–0.268	0.490	0.193	0.572	0.571
ICT	–0.002	–0.134	0.129	0.067	–0.037	0.970
AD	0.393	–0.275	1.062	0.341	1.152	0.257
MFT	0.022	–0.351	0.397	0.191	0.119	0.905
MAIN	–0.057	–0.398	0.283	0.173	–0.330	0.743

< .10 = . ; < .05 = * ; < .01 = ** ; < .001 = ***.

Data accessibility statement

The anonymised data files and an R-script with the presented analyses are available at OSF https://osf.io/zbfuq/.

Ethics and consent

The MINT project was conducted in accordance with the regulations of The Swedish Data Protection Authority and the Ethical Review Board at Karolinska Institutet (Dnr 2011/955-31/1; Dnr 2020-03578) and The Personal Data Act (1998: 204) and the Act concerning the Ethical Review of Research Involving Humans (2003: 460).

Acknowledgements

The study is based on the MINT corpus and would not have been possible without the countless hours spent on data collection, annotation, analyses, and administration within the MINT project, under the direction of Tove Nilsson Gerholm. The first author wants to extend his gratitude towards the project and all who have contributed to it.

The authors would also like to extend their gratitude towards Danielle Matthews and Kirsten Abbot-Smith for their invaluable contributions through their collaboration with the first author, preceding this study.

Competing interests

The authors have no competing interests to declare.

Author contributions

David Pagmar: Conceptualisation (equal), Data curation (equal), Formal Analysis, Investigation (equal), Methodology (equal), Software, Visualisation, Writing – original draft (lead), Writing – review & editing (equal). Caroline Arvidsson: Data curation (equal), Investigation (equal). Tove Nilsson Gerholm: Data curation (equal), Investigation (equal), Project administration, Project leader, Resources (lead), Writing – review & editing (support). Julia Uddén: Conceptualisation (equal), Methodology (equal), Resources (supportive), Supervision, Writing – original draft (support), Writing – review & editing (equal).

References

Abbot-Smith, K., Matthews, D., Malkin, L., & Nice, J. (2021). Cognitive correlates of conversational contingency in autistic and neuro-typical children. DOI: https://doi.org/10.17605/OSF.IO/Q7WA4

Ahlström, L., & Ljungman, H. (2011). Åldersreferenser för Peabody Picture Vocabulary Test IV på svenska för flerspråkiga barn i skolår 4. *Unpublished master thesis, Karolinska Institute, Stockholm, Sweden*.

Andrés-Roqueta, C., & Katsos, N. (2017). The contribution of grammar, vocabulary and theory of mind in pragmatic language competence in children with autistic spectrum disorders. *Frontiers in Psychology*, *8*, 996. DOI: https://doi.org/10.3389/fpsyg.2017.00996

Arvidsson, C., Pagmar, D., & Uddén, J. (2022). When did you stop speaking to yourself? Age-related differences in adolescents' world knowledge-based audience design. *Royal Society Open Science*, *9*(11), 220305. DOI: https://doi.org/10.1098/rsos.220305

Bacso, S. A., & Nilsen, E. S. (2017). What's that you're saying? Children with better executive functioning produce and repair communication more effectively. *Journal of Cognition and Development*, *18*(4), 441–464. DOI: https://doi.org/10.1080/15248372.2017.1336438

Becker, J. A. (1994). Pragmatic socialization: Parental input to preschoolers. *Discourse Processes*, *17*(1), 131–148. DOI: https://doi.org/10.1080/01638539409544862

Bendtz, K., Ericsson, S., Schneider, J., Borg, J., Bašnáková, J., & Uddén, J. (2022). Individual differences in indirect speech act processing found outside the language network. *Neurobiology of Language*, *3*(2), 287–317. DOI: https://doi.org/10.1162/nol_a_00066

Bernard, S., & Deleau, M. (2007). Conversational perspective-taking and false belief attribution: A longitudinal study. *British Journal of Developmental Psychology*, *25*(3), 443–460. DOI: https://doi.org/10.1348/026151006X171451

Bishop, D. V. (1998). Development of the Children's Communication Checklist (CCC): a method for assessing qualitative aspects of communicative impairment in children. *Journal of child psychology*

and psychiatry, and allied disciplines, *39*(6), 879–891. DOI: https://doi.org/10.1111/1469-7610.00388

Blain-Brière, B., Bouchard, C., & Bigras, N. (2014). The role of executive functions in the pragmatic skills of children age 4–5. *Frontiers in psychology*, *5*, 240. DOI: https://doi.org/10.3389/fpsyg.2014.00240

Bloom, L., Rocissano, L., & Wootten, J. (1982). Adult-Child Discourse: Developmental Interaction Between Information Processing and Linguistic Knowledge. Cognitive Psychology, *8*, 521–551. DOI: https://doi.org/10.1016/0010-0285(76)90017-7

Bohnacker, U., & Gagarina, N. (Eds.). (2020). *Developing narrative comprehension: Multilingual assessment instrument for narratives* (Vol. 61). John Benjamins Publishing Company. DOI: https://doi.org/10.1075/sibil.61

Bornstein, M. H., Hahn, C. S., & Putnick, D. L. (2016). Long-term stability of core language skill in children with contrasting language skills. *Developmental Psychology*, *52*(5), 704. DOI: https://doi.org/10.1037/dev0000111

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental neuropsychology*, *26*(2), 571–593. DOI: https://doi.org/10.1207/s15326942dn2602_3

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delays. *Autism*, *2*(4), 325–344. DOI: https://doi.org/10.1177/1362361398024002

Clark, E. 1993. The lexicon in acquisition. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511554377

Dahlgren, S., & Sandberg, A. D. (2008). Referential communication in children with autism spectrum disorder. *Autism*, *12*(4), 335–348. DOI: https://doi.org/10.1177/1362361308091648

De Rosnay, M., Fink, E., Begeer, S., Slaughter, V., & Peterson, C. (2014). Talking theory of mind talk: Young school-aged children's everyday conversation and understanding of mind and emotion. *Journal of child language*, *41*(5), 1179–1193. DOI: https://doi.org/10.1017/S0305000913000433

Dunn, L. M., & Dunn IV, L. M. (2007). PPVT-IV Test kit for form IV. *Circle Pines, MN: AGS*. DOI: https://doi.org/10.1037/t15145-000

ELAN (Version 6.4) [Computer software]. (2022). Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i–185. PMID: 7845413. DOI: https://doi.org/10.2307/1166093

Forrester, M. A., & Cherington, S. M. (2009). The development of other-related conversational conducts: A case study of conversational repair during the early years. First language, *29*(2), 166–191. DOI: https://doi.org/10.1177/0142723708094452

Foppolo, F., Mazzaggio, G., Panzeri, F., & Surian, L. (2021). Scalar and ad-hoc pragmatic inferences in children: Guess which one is easier. *Journal of Child Language*, *48*(2), 350–372. DOI: https://doi.org/10.1017/S030500092000032X

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at Head Start. *Developmental psychology*, *47*, 404–416. DOI: https://doi.org/10.1037/a0021065

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., … & Walters, J. (2015). Assessment of narrative abilities in bilingual children. *Assessing multilingual children: Disentangling bilingualism from language impairment*, *241*, 274. urn:nbn:se:uu:diva-279888. DOI: https://doi.org/10.21832/9781783093137-011

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics*, *63*, 20. DOI: https://doi.org/10.21248/zaspil.63.2019.516

Gerholm, T. (2018). *Conventions for annotation and transcription of the MINT-project: Modulating child language acquisition through parent-child interaction, MAW: 2011.007*. Stockholm University.

Gerholm, T., Kallioinen, P., Tonér, S., Frankenberg, S., Kjällander, S., Palmer, A., & Lenz-Taguchi, H. (2019). A randomized controlled trial to examine the effect of two teaching methods on preschool children's language and communication, executive functions, socioemotional comprehension, and early math skills. *BMC psychology*, *7*(1), 1–28. DOI: https://doi.org/10.1186/s40359-019-0325-9

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill. DOI: https://doi.org/10.1163/9789004368811_003

Hale, C. M., & Tager-Flusberg, H. (2005). Social communication in children with autism: The relationship between theory of mind and discourse development. *Autism*, *9*(2), 157–178. DOI: https://doi.org/10.1177/1362361305051395

Hazen, N. L., & Black, B. (1989). Preschool peer communication skills: The role of social status and interaction context. *Child development*, 867–876. DOI: https://doi.org/10.2307/1131028

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. DOI: https://doi.org/10.1111/desc.12536

Hoff-Ginsberg, E. (1998). The relation of birth order and socio-economic status to children's language experience and language development. *Applied Psycholinguistics*, *19*, 603–629. DOI: https://doi.org/10.1017/S0142716400010389

Huizinga, M., Dolan, C. V., & Van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036. DOI: https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Katsos, N. (2014). Scalar implicature. *Pragmatic Development in First Language Acquisition*, *10*, 183–197. DOI: https://doi.org/10.1075/tilar.10.11kat

Kissine, M. (2021). Autism, constructionism, and nativism. *Language 97*(3), e139–e160. DOI: https://doi.org/10.1353/lan.2021.0055

Küntay, A. C., & Ozyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal?. *Journal of Child Language*, *33*(2), 303–320. DOI: https://doi.org/10.1017/S0305000906007380

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, *32*(1), 69–92. DOI: https://doi.org/10.1017/S0142716410000299

Lehto, J. (1996). Are executive function tests dependent on working memory capacity?. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 29–50. DOI: https://doi.org/10.1080/713755616

McCaw, Z. (2022). RNOmni: Rank Normal Transformation Omnibus Test. Version 1.0.1.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732–749. DOI: https://doi.org/10.1111/j.1467-8624.2008.01154.x

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition, and executive functions. *Language Learning and Development*, *14*(3), 186–223. DOI: https://doi.org/10.1080/15475441.2018.1455584

Mognon, I., Sprenger, S. A., Kuijper, S., & Hendriks, P. (2021). Complex inferential processes are needed for implicature comprehension, but not for implicature production. *Frontiers in Psychology*, *11*, 556667. DOI: https://doi.org/10.3389/fpsyg.2020.556667

Nadig, A., Lee, I., Singh, L., Bosshart, K., & Ozonoff, S. (2010). How does the topic of conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. *Neuropsychologia*, *48*(9), 2730–2739. DOI: https://doi.org/10.1016/j.neuropsychologia.2010.05.020

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive psychology*, *58*(2), 220–249. DOI: https://doi.org/10.1016/j.cogpsych.2008.07.002

Noveck, I. A. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188. DOI: https://doi.org/10.1016/S0010-0277(00)00114-1

Pagmar, D. (2023). 2. Background. In *The development of conversational contingency and selected pragmatic abilities* (pp 5–22). *Doctoral dissertation*. Stockholm University.

Pagmar, D., Abbot-Smith, K., & Matthews, D. (2022). Predictors of children's conversational contingency. *Language Development Research*, *2*(1), 139–179. DOI: https://doi.org/10.34842/2022-511

Pagmar, D., Arvidsson, C., & Uddén, J. (2021). Testing for audience design: A short report. Preprint at PsyArXiv. DOI: https://doi.org/10.31234/osf.io/3d7ja

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition: A Journal of Developmental Linguistics, 12*(1), 71–82. DOI: https://doi.org/10.1207/s15327817la1201_3

Peterson, B. G., & Carl, P. (2020). PerformanceAnalytics: Econometric tools for performance and risk analysis. Version 2.0.4

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, https://www.R-project.org/.

Song, S., Su, M., Kang, C., Liu, H., Zhang, Y., McBride-Chang, C., ... & Shu, H. (2015). Tracing children's vocabulary development from preschool through the school-age years: An 8-year longitudinal study. *Developmental Science*, *18*(1), 119–131. DOI: https://doi.org/10.1111/desc.12190

Stratta, P., Daneluzzo, E., Prosperini, P., Bustini, M., Mattei, P., & Rossi, A. (1997). Is Wisconsin Card Sorting Test performance related to 'working memory' capacity? *Schizophrenia Research*, *27*(1), 11–19. DOI: https://doi.org/10.1016/S0920-9964(97)00090-X

Tager-Flusberg, H., & Anderson, M. (1991). The development of contingent discourse ability in autistic children. *Journal of Child Psychology and Psychiatry*, *32*(7), 1123–1134. DOI: https://doi.org/10.1111/j.1469-7610.1991.tb00353.x

Tonér, S., & Gerholm, T. N. (2021). Links between language and executive functions in Swedish preschool children: A pilot study. *Applied Psycholinguistics*, *42*(1), 207–241. DOI: https://doi.org/10.1017/S0142716420000703

Richard, D. M., & Rouder, J. N. (2021) *BayesFactor: Computation of Bayes Factors for Common Designs.* Version 0.9.12–4.3.

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, *83*(2), 508–525. DOI: https://doi.org/10.1111/j.1467-8624.2011.01710.x

Van Eylen, L., Boets, B., Steyaert, J., Evers, K., Wagemans, J., & Noens, I. (2011). Cognitive flexibility in autism spectrum disorder: Explaining the inconsistencies? *Research in Autism Spectrum Disorders*, *5*(4), 1390–1401. DOI: https://doi.org/10.1016/j.rasd.2011.01.025

Wanska, S. K., & Bedrosian, J. L. (1985). Conversational Structure and Topic Performance Mother-Child Interaction. *Journal of Speech, Language, and Hearing Research*, *28*(4), 579–584. DOI: https://doi.org/10.1044/jshr.2804.579

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, *44*, 575–587. DOI: https://doi.org/10.1037/0012-1649.44.2.575

Wilson, A. C., & Bishop, D. V. (2019). "If you catch my drift...": Ability to infer implied meaning is distinct from vocabulary and grammar skills. *Wellcome open research*, *4*. DOI: https://doi.org/10.12688/wellcomeopenres.15210.3

Wilson, A. C., & Bishop, D. V. (2022). A novel online assessment of pragmatic and core language skills: An attempt to tease apart language domains in children. *Journal of Child Language*, *49*(1), 38–59. DOI: https://doi.org/10.1017/S0305000920000690

Yliherva, A., Loukusa, S., Väisänen, R., Pyper, A., & Moilanen, I. (2009). Development of communication skills in Finnish pre-school children examined by the Children's Communication Checklist (CCC). *Child Language Teaching & Therapy*, *25*, 235–249. DOI: https://doi.org/10.1177/0265659009102978