

Pragmatic and core language skills

**A novel online assessment of pragmatic and core language skills: An attempt to tease apart
language domains in children**

WILSON, Alexander C.

BISHOP, Dorothy V. M.

Department of Experimental Psychology, University of Oxford

alexander.wilson2@psy.ox.ac.uk

Key words

Language development; pragmatics; assessment

Abstract

It remains unclear whether pragmatic language skills and core language skills (grammar and vocabulary) are distinct language domains. The present work aimed to tease apart these domains using a novel online assessment battery administered to almost 400 children aged 7 to 13 years. Confirmatory factor analysis indicated that pragmatic and core language domains could be measured separately, but that both domains were highly related ($r = .79$). However, zero-order correlations between pragmatic tests were quite small, indicating that task-specific skills played an important role in performance, and follow-up exploratory factor analysis suggested that pragmatics might be best understood as a family of skills rather than a domain. This means that these different pragmatic skills may have different cognitive underpinnings and also need to be assessed separately. However, our overall results supported the idea that pragmatic and core aspects of language are closely related during development, with one area scaffolding development in the other.

Pragmatic and core language skills

While vocabulary acquisition and grammatical skills are key to language development, they are not sufficient by themselves for understanding language. Alongside decoding processes (which include vocabulary and grammatical skills), language comprehension also relies on pragmatic processing, whereby we produce an elaborate understanding of what the speaker intended to communicate in the context (Ariel, 2010). Relevance Theory is an influential model of communication that makes this distinction between decoding and pragmatics (Sperber & Wilson, 1986). It recognises that all uses of language involve some level of incompleteness or ambiguity, and we must therefore use the broader communication context and the assumption that utterances are tailored relevantly to their context (i.e. the principle of optimal relevance) to infer the full extent of the speaker's intended meaning. As an example, consider the utterance "it's quite cold". Here, the speaker could be implying they'd like to go inside, warning a child not to go in the water, or complaining about their cold soup! Pragmatic processing would allow us to settle on the most likely meaning, and this would of course depend on the context. As noted by Cummings (2007, pp. 425–426), pragmatics is "a process of reasoning based on features of context".

Some uncertainty exists about the relationship between pragmatics and what we term "core language skills" (grammar and vocabulary). On the one hand, there is the clinical intuition that some individuals may struggle specifically with using and understanding language in social situations. This profile has been conceptualised as "pragmatic language impairment" (Bishop, 1998) and has recently been formalised as the diagnosis "social (pragmatic) communication disorder" in DSM-5 (American Psychiatric Association, 2013). These labels suggest that an impairment in using and understanding language in context may co-exist with typical (or indeed

Pragmatic and core language skills

advanced) vocabulary and grammar development (i.e. typical core language development). This has the broader implication that these two sets of skills (communication in context and core language) are separate domains. On the other hand, it has been surprisingly hard to provide empirical evidence for this using objective tests, and research has generally indicated that tests of pragmatics and core language are highly correlated (Matthews, Biney, & Abbot-Smith, 2018).

In recent work, we suggested that these high correlations may be due to a lack of assessments that specifically measure pragmatic processing (Wilson & Bishop, 2019). Accordingly, we devised a pragmatic test focusing on particularised conversational implicature, designed to target pragmatic processing while minimising other linguistic demands. Particularised conversational implicature involves meaning that can only be inferred from the context, as in the following exchange: “Did you hear what the police said?” “There were lots of trains going past.” On its own, the second utterance wouldn’t bring to mind the police; it is only in the context of the question that we pick up an implied meaning relating to the police, i.e. that the speaker didn’t hear what they said. This kind of meaning depends on a sensitivity to the context of communication. We proposed, therefore, that a test of particularised conversational implicature would represent a valid test of pragmatic comprehension (that is, inferring meaning from context), and would be a relatively specific test of this skill if vocabulary and grammatical structure were kept simple. In a sample of 120 adults, we found that scores on our implicature comprehension test were only weakly related to vocabulary and grammar ability, supporting the view that core language and pragmatics are somewhat separable. Our purpose in this study was to test whether this is also true in children whose language skills are still developing.

Implied meaning and core language in children

As with the adult sample, we expected to find some relationship between implicature comprehension and core language in children. Greater vocabulary and grammatical skills are likely to mean a child will more easily process the “explicit” content of an utterance, which is therefore more fully available for “reading between the lines” to grasp what the speaker implicitly means in context. Indeed, implicature comprehension and core language may be more interrelated in children than in adults. In adults, the linguistic code may be stored in a more self-encapsulated way, as it represents “crystallized” knowledge“. But children, of course, enter the linguistic environment without a lexicon, and during language development they may depend on more continual interaction between pragmatic and core language skills, so there may be few situations in which they rely on one domain in relative exclusion of the other. It is likely that this interaction runs in both directions: from pragmatics to core language, and vice versa. For instance, a child with well-developed core language may understand more of what they hear, giving more opportunity for learning to infer from context. In a reciprocal way, a child who is more sensitive to the importance of context in communication may be more resourceful in interpreting unfamiliar language, thereby accelerating their acquisition of core language skills. However, it is possible that pragmatics and core language show a similar relationship in children as adults, if these different processes are “modular”; certainly, pragmatic processing of communicative stimuli is seen as modular by Relevance Theory (Carston, 2002).

However, existing studies would tend to support the view that there is interaction between domains. In approximately 95 typically developing children aged 4 to 12 years, a composite score for sentence comprehension and vocabulary skills was moderately correlated ($r = .46$) with

Pragmatic and core language skills

implicature comprehension (including manner, scalar and metaphor implicatures) (Antoniou & Katsos, 2017). Similarly, research indicates that while children with communication difficulties perform lower on tests of figurative/nonliteral/ambiguous language, this is typically at a level that directly relates to their core language skills rather than reflecting a specific pragmatic problem (for reviews, see Andrés-Roqueta & Katsos, 2017; Kalandadze, Norbury, Nærland, & Næss, 2018). For an alternative view, see the review by Loukusa and Moilanen (2009) which emphasises the importance of other aspects of cognition, including theory of mind and central coherence, in making inferences in a range of language tasks, although it is arguable whether enough attention is given in the review to the role played by core language abilities. It is worth noting that core language demand is not usually controlled in language tasks involving inferencing and implicature, making it difficult to assess the extent to which processing “explicit” and more implicit language relies on different skills. It is also rare to find studies that present language in conversational contexts – i.e. the kind of contexts where fast-paced, context-dependent pragmatic processing might be most important. Even with these provisos, though, it is unreasonable to suggest that a “pure” test of language pragmatics, if such a thing were possible, would be entirely distinct from core language.

Social-pragmatic and core language abilities in children

A test of particularised conversational implicature offers just one measure of pragmatic processing, whereas a more optimal method of measuring an individual’s pragmatic ability would involve administering a range of pragmatic subtests before extracting a general factor of pragmatic ability from them. In Wilson and Bishop (2019), we used factor analysis to test whether implicature comprehension clustered as part of a more general social cognition factor,

Pragmatic and core language skills

distinct from core language skills, in adults. In line with theoretical predictions of Relevance Theory (Sperber & Wilson, 1986), we found that scores on the test battery fitted a model with two factors, a “core language” and “social understanding/inferential processing” factor, with implicature comprehension loading on the second factor. This finding comes with three qualifications. Firstly, the tests under the second factor showed relatively weak relationships with each other, suggesting that specific skills may be as important as a general ability for performance on these tests of social understanding/inferential processing, at least in a general population sample. Secondly, there was evidence that the factor originally labelled as “social understanding” was not specifically social in nature after all, but was better described as domain-general inferential processing. Finally, there was a moderate correlation between the two factors ($r = .55$), indicating that language skills and social understanding/inferential processing are quite closely related.

In the work reported here, we administered a test battery to a sample of children to see whether a pragmatic language factor emerged as distinct from core language. To complement the Implicature Comprehension Test, we devised three new pragmatic language tests. The first asked participants to draw inferences across sections of spoken text, the second involved detecting violations of pragmatic principles, the third required sensitivity to the social intent of conversational overtures. For the text-based inference task, participants needed to process meanings that were not fully stated, as with the implicature task, although this was in the context of a short story. For the pragmatic violations task, we devised short conversational interchanges in which a character violated the principle of optimal relevance; either they gave insufficient or redundant information in response to a question, or they drifted from the topic. As well as being

Pragmatic and core language skills

motivated by Relevance Theory, these categories of violation represent a range of pragmatic infelicities made by children with and without language and communication difficulties (Bishop & Adams, 1989). The other new pragmatic task involved conversational overtures; a character produces utterances such as “I had a really bad journey today” or “I’m going to close the windows”, and the participant needs to decide if the character is trying to start a conversation with the other character or just providing a brief comment. As such, this is a test requiring us to read communicative intentions based on a sentence in a social context. This task is clinically relevant as responses to conversational overtures are evaluated in the ADOS-2 (Lord et al., 2012), the widely used tool for assessing communication skills in autistic people. Alongside the implicature test and these three new pragmatic tasks, tests of receptive grammar and vocabulary were administered as measures of core language.

We had two hypotheses:

- (1) Comprehension of conversational implicature would only be weakly related to grammar and vocabulary skills after controlling for age.
- (2) In factor analysis, a model with pragmatic and core language skills as separable factors would provide a better fit to the data than a single factor model.

We also planned some follow-up analyses. In the case of pragmatics and core language skills being related, this may be because they are both reflections of general cognitive ability rather than because they are both linguistic, so it was important to assess whether any relationship was accounted for by nonverbal ability. We also aimed to assess the relationship between the language factors and sex and age.

Methods

This study was a partial replication of the work reported in Wilson and Bishop (2019), which involved a factor analysis of different language and social/communication skills in a general population sample of adults; the present study tested the factor structure of a similar set of tasks in typically-developing children. We report below how we determined the sample size, all data exclusions, all manipulations, and all measures in the study. The study was granted ethical clearance in November 2018 by the Medical Sciences Interdivisional Research Ethics Committee at the University of Oxford: Ref. R59912/RE001.

Power calculation

To determine the minimum sample size, we used an adapted form of the power calculation from Wilson and Bishop (2019). This method involved creating simulated datasets with known factor structure, and repeatedly running the analysis with different sample sizes to establish the sample size at which the factor structure could be recovered. Using the R package *simsem* (Jorgensen, Pornprasertmanit, Miller, & Schoemann, 2018), we simulated datasets conforming to a two-factor model at the population level, with two indicators loading on one factor and four indicators loading on the other factor. Factor loadings in the population model were set to .7 and factor correlation to .6. In 10,000 simulations of samples of 120 individuals, 9980 datasets showed a significantly better fit to a two-factor compared to a one-factor model, when compared using a chi-square test with the alpha level set to .05. Thus this sample size gives 99.8% power to detect the improved fit of a two-factor model at the .05 level.

Participants

We aimed to recruit at least 120 child participants aged between 7;0 and 13;11 years, by inviting all the year 3 to 8 children in several schools to take part. We used an opt-out approach to recruitment; schools were asked to circulate information to parents/carers two weeks before research sessions, and if parents/carers preferred their child not to be involved, they returned an opt-out form to school. Children who were happy to take part in the research were asked to complete an assent form, and we informally checked ongoing assent throughout the research sessions. For any child preferring not to take part, teachers discreetly arranged for them to complete other computer-based activities that were part of the normal week-to-week learning. The data from four children identified by their teachers as having an autism diagnosis were excluded from the analysis.

To maintain anonymity of the children involved in our project, we did not collect personally-identifying information. Teachers were simply asked against each child's research ID to indicate their year group (i.e. years 3 to 8), gender, whether they were diagnosed with autism, and whether they spoke English as an additional language.

The sample consisted of 390 children. See Table 1 for details of the sample, including year group, sex, English-speaking status and missing data. Children speaking English as an additional language tended to perform as well on the tasks as native speakers; it was only on vocabulary that a small advantage emerged for native speakers, who had a mean of 28.78 compared to 27.39, $t(75.89) = 2.26$, $p = .027$. Due to the practicalities of working in busy schools over several testing sessions, and IT difficulties, there is some missing data. Of the primary-age children (i.e. years 3 to 6, which comprise 7 to 11-year-olds), just over 30% did not complete at

least one of the tests. There is considerable missing data for the youngest children; this was not a consequence of any particular challenges these children experienced with testing but instead due to IT difficulties that occurred as a matter of coincidence with two year 3 classes. As for the secondary-age children (years 7 and 8, comprising 11 to 13-year-olds), we did not set out to collect full data from any one child, owing to limited timetabling opportunities. With this group, the order of tasks was randomised and children completed as many tasks as possible.

Table 1

Sample Characteristics. Teachers reported the sex of children in their classes, as well the number who spoke English as an additional language (EAL). Here we also indicate the number of children for whom there are complete data.

	N children	Girls	Boys	Sex not reported	EAL	Complete data
Year 3	54	29	25	0	15	21
Year 4	77	23	27	27	16	54
Year 5	83	42	41	0	8	72
Year 6	81	32	25	24	13	56
Year 7	51	24	27	0	0	0
Year 8	44	24	20	0	1	0

Procedure

For the primary school classes (years 3 to 6), two group-based testing sessions were arranged in school with each class. Children completed a battery of newly devised tasks during each session. Sessions lasted not longer than 50 minutes. Children worked on a sequence of computerised tasks, listening to audio through headphones and supervised by the first author and a member of the school teaching staff. The tasks were supported by Gorilla (<https://gorilla.sc/>), a cloud-based tool for collecting data in the behavioural sciences. In the first session, children completed the Implicature Comprehension Test, Receptive Vocabulary task, Pragmatic

Pragmatic and core language skills

Violations and Animal Matrices, in that order. In the second session, children completed the Social Overtures, Receptive Grammar task, and the Children's Test of Local Textual Inference two weeks after the first session. During this second session, children also repeated one task from the first session; this task was randomly assigned to each child, and was administered to assess test-retest reliability.

For the secondary school classes (years 7 and 8), children took part in one group-based testing session lasting 50 minutes. The order of tasks was randomised, and children completed as many tasks as they could manage in the time.

Measures

1. *Pragmatic language tasks* (In the confirmatory factor analysis, these tasks were set to load on the “pragmatic language factor”.)

Implicature Comprehension Test (Wilson & Bishop, 2019). Participants watch a series of cartoon videos, in each of which two characters produce a short utterance one after the other. Together the utterances form a conversational adjacency pair; in most cases, this is a question and answer. After this dialogue, participants hear a comprehension question, and they give a yes-no-don't know response by clicking buttons on the screen.

For 36 items, participants need to process implied meaning to answer the question, as the second character provides an indirect response to the first character. An example item includes:

Character 1: “Could you hear what the police said?” Character 2: “There were lots of trains going past.” Comprehension Question: “Do you think she heard what the police said?” Correct Answer: “No.”

Pragmatic and core language skills

There are also 10 items where the answer is more explicit; these serve as positive control items.

An example item includes:

Character 1: “Did you see the policemen earlier on?” Character 2: “I saw them standing on the platform.” Comprehension Question: “Do you think he saw the policemen?”

Correct Answer: “Yes.”

From these items, there were two measured variables: sum of implicature items correctly answered (out of 33; three items were excluded, as mentioned in Results) and sum of explicit-response items correctly answered (out of 10).

Children’s Test of Local Textual Inference. This is based on the adult version of this test (Wilson & Bishop, 2019). Participants hear two brief sections of a short story (about 90 words per part). After each section, they hear ten questions and four possible answers for each one. “We don’t know” is an answer option for every question, and is the correct answer to four questions. Participants click the correct option on the screen. As well as auditory presentation of all materials, everything is shown in text-based form on the screen. Participants are informed at the start that the short story sections will remain on the screen while they are answering questions about that section. Participants need to make inferences based on the short story to answer the questions. The short story starts as follows: “Unfortunately, the family couldn’t go swimming. The sea was rougher and colder than expected. Instead, Billy spent the whole morning playing a ballgame with his sister, Susie.” An example question is: “What had Billy planned to do?” Participants chose their answer from the following options: “play a ballgame”, “go swimming”,

Pragmatic and core language skills

“walk along the sea”, and “we don’t know”. There was one measured variable: the sum of items correctly answered (out of 20).

Pragmatic Violations. Participants watch a sequence of 35 videos, in which one character produces a question and a second character answers the question. In 21 videos, the second character violates a pragmatic norm; in the others, the answer is appropriate. Here is an example of a good item: “Why did the dog bark?” “Someone came to the door. They made a lot of noise. It scared the dog.” There are four types of pragmatic violation: redundancy ($n = 6$), missing information ($n = 6$), topic drift ($n = 6$) and unresponsiveness ($n = 3$). The unresponsive items are designed to be trivially easy, as the answer is simply “er”. In the redundant items, the second character states information that can be readily assumed from preceding sentences or world knowledge; e.g. “Where did you go yesterday?” “Yesterday I went to the swimming pool. I wanted to go swimming yesterday. I had to take my towel with me.” In the topic drift items, the second character shifts the topic under discussion in a tangential manner; e.g. “Where is the dog?” “I let the dog outside into the garden. I didn’t give it any toys. It chewed up a teddy bear last week.” In the missing information items, the second character does not provide the full context/background to their answer; e.g. “Did you go to the cinema?” “I missed the film. He said sorry.” Individuals are asked whether they think the second character gave a “good answer”, and need to rate how good the answer was by placing a mark on an analogue scale showing a sad face at one end and a happy face at the other. The mark can be placed at seven increments along the line. Although numbers are not shown on the scale, scores between 1 and 7 are directly assigned for participants’ responses according to where they put the mark on the line. The measured variable is the total number of pragmatic violations correctly identified. To calculate a

Pragmatic and core language skills

participant's total score, we first identified the score given to the 21 items the participant thought were the worst (as 21 items were devised to include violations). Items with a pragmatic violation given a score below that level were given one point. For items with a pragmatic violation that were given a score at that level, we gave one point divided by the number of items given that score (to allow for tied items). During piloting with 30 non-autistic adults, three of the items did not correlate with the others (one missing information item and two redundancy items). To improve the reliability of the measure, these items were excluded from the total score. Cronbach's alpha calculated from the adult pilot data was .81 [95% CIs .70, .89].

Social Overtures. Participants hear 24 utterances spoken by a character to a conversational partner. Twelve are social overtures that attempt to engage the partner in a conversation (e.g. "I can't believe what happened today.") and twelve are not conversational bids (e.g. "I'm going to have a shower now."). Participants listen to instructions explaining that "There are different reasons why we say things to other people. Sometimes, we want to start a conversation. We want the other person to ask us questions and say lots of things to us. Other times we just want to tell the other person something very quickly. We don't always want to start a long conversation." They are then asked for each sentence whether the speaker wants a conversation or not, and to indicate their answer by clicking yes-no buttons. There was one measured variable: the sum of items correctly identified as a social overture or not (out of 23; one item was excluded, as mentioned in Results).

2. Core language tasks (In the factor analysis, these tasks were set to load on the "core language" factor. As tasks were presented in an online format, it was not possible to use current standardised measures, as these have not been validated for this format.)

Receptive Vocabulary. This includes 40 items in which participants choose which of four pictures is related to a word. Participants hear a sequence of 40 words and for each word, they are presented with four pictures on the screen. They are asked to “choose which picture goes best with the word”. The words include nouns, verbs and adjectives, and vary in approximate age of acquisition from 5 to 12, with similar numbers of easy and harder words; two experienced teachers independently rated the ages at which they would expect 50% and 90% of children in a typical class to be familiar with the word. There was one measured variable: the sum of items correctly answered (out of 39; one item was excluded, as mentioned in Results).

Receptive Grammar. Participants listen to sentences and decide if they are grammatical. They hear the following instructions: “Some of the sentences will sound good, but some of the sentences will sound bad. There might be a missing word. Or the wrong word might be used. Or the order of the words might be weird. If the sentence is good, click the green tick. If the sentence is bad, click the red cross”. There are 50 items: in 4 sentences the words are in a random order and should be easily rejected, 20 items are taken from McDonald (2006) and showed high accuracy in primary school children, and 26 items are a subset of our adult version of this test (Wilson and Bishop, 2019); these latter items were chosen on the basis of high accuracy and high item-total correlations. Excluding the 4 randomly ordered sentences, 23 items do not follow typical syntax or use incorrect word forms (e.g. incorrect tenses) and 23 follow typical English grammar. Examples of incorrect items include: “The teacher told the story the children” and “I went out after I have eaten dinner”. There was one measured variable: the sum of items currently answered (out of 50).

3. General cognitive ability

Animal Matrices. This non-verbal reasoning task is an adapted version of the Animalogica multiple choice test (Stevenson, Heiser, & Resing, 2016). There are 18 items. Each item is a 2x2 matrix presented on the computer screen. In three of the boxes of each matrix, there are cartoon pictures of animals, and the fourth box is empty. The animals in the three boxes vary along six dimensions: species, colour, size, number, direction faced, and position in the box. There are systematic relationships between the three animals, and participants need to deduce which of five options fits in the empty box. For example, the top two boxes may show red lions, one big and one small, and the bottom left box may show a big yellow horse; the correct option to fill the empty box would be a small yellow horse. There was one measured variable: the sum of items correctly answered (out of 16; two items was excluded, as mentioned in Results).

4. Standardised cognitive measures

Vocabulary and Matrix Reasoning. 20 children aged 9 to 10 completed these two subtests from the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) so that the validity of the novel tasks could be evaluated in comparison to these standardised assessments.

Data Analysis

Data were analysed in R (R Core Team, 2017). Data and scripts are accessible on OSF (<https://osf.io/xya3j/>).

As the tests were novel, we began by assessing item functioning and test reliability to establish the quality of the tests. We used classical test theory (CTT) and item response theory (IRT; Embretson & Reise, 2000) approaches, implemented with R packages psych (Revelle, 2017) and mirt (Chalmers, 2012). For each item in each test, we inspected accuracy and the

Pragmatic and core language skills

correlation between item-accuracy and total-accuracy on the test with that item excluded. Items were identified as poor if they had low accuracy and a low item-total correlation. We also inspected item characteristic curves (ICCs) produced using IRT analysis, which determines difficulty and discrimination parameters for each item. ICCs are useful in identifying poor items, as they show the probability that an individual of a certain ability scores at a particular level on an item; low flat curves indicate items that are ambiguous or too difficult for the participant group, with no consensus answer at any point along the ability spectrum. We excluded any items showing this pattern. We then computed the reliability of each of the tests using CTT coefficients, Cronbach's alpha (with 95% confidence intervals) and Revelle's beta. Revelle's beta is the worst split-half reliability – i.e. the correlation between the halves of a test when it is split in such a way that the correlation is minimised (Zinbarg, Revelle, Yovel, & Li, 2005). We used beta to give an estimation of general factor saturation that removes the influence of any subgroup factors. Where there is a large discrepancy between alpha and beta, the measure is heterogeneous in its content. As an indication of the unidimensionality of the tests, we report the root mean square error of estimation (RMSEA) for a unidimensional IRT model for each test. Test-retest reliability was also assessed where tests had been administered twice by computing correlations between the two administrations. As noted in the Procedure, the primary school children (years 3 to 6) were randomly assigned one task from the first session (Implicature Comprehension Test, Pragmatic Violations, Receptive Vocabulary or Animal Matrices) to complete again during the second testing session two weeks later.

To test the relationship between implicature comprehension and core language skills, we ran a hierarchical multiple regression using data from individuals who completed the three tests

Pragmatic and core language skills

relevant to this analysis. In the first stage of the regression, year group was included as a predictor of scores on the implicature items of the Implicature Comprehension Test; in the second stage, Receptive Vocabulary and Receptive Grammar were included as predictors. In a third stage, scores on the Children's Test of Local Textual Coherence were added as a predictor, to assess whether processing implied meaning in narratives and conversation showed a specific relationship, even controlling for the general role played by core language skills. Residual plots were inspected to check whether assumptions of regression were met (Altman & Krzywinski, 2016). Adjusted R-squared was the effect size, with 95% confidence intervals estimated through bootstrapping (using R package *boot*; Canty & Ripley, 2017).

For the second hypothesis, we used confirmatory factor analysis to test whether implicature comprehension forms part of a “pragmatic language” factor that is distinct from “core language”. First, we inspected the data for normality and multivariate outliers using R package *MVN* (Korkmaz S, Goksuluk, & Zararsiz, 2014). Multivariate outliers were defined as individuals whose adjusted Mahalanobis' distance was above the 97.5th percentile of the chi-distribution. Maximum likelihood estimation was used for the factor analysis. As this is based on the multivariate normal distribution, it was important to check whether our data departed from this distribution (Flora, Labrish, & Chalmers, 2012), although given that robust estimation was used in the factor analysis, assumptions of normality were relaxed (Finney & DiStefano, 2013). Nonetheless, in the case of multivariate outliers being present, we planned to transform variables using the Tukey ladder of power transformations using R package *rcompanion* (Mangiafico, 2018) to reduce skew, and then test for multivariate outliers again. Remaining outliers would be excluded. The sensitivity of the analysis to any data transformation or outlier exclusion was

Pragmatic and core language skills

evaluated by comparing results based on transformed and non-transformed data with and without outliers included.

For the confirmatory factor analysis, we used R package lavaan (Rosseel, 2012), and semtools (Jorgensen et al., 2019) for the visual representations of our analysis. We specified a two-factor correlated-traits model with a “core language” factor and a “pragmatic language” factor. Two indicators were set to load on the “core language” factor: totals for Receptive Grammar and Receptive Vocabulary. Four indicators were set to load on the “pragmatic language” factor: totals for the Implicature Comprehension Test, Children’s Test of Local Textual Inference, Pragmatic Violations, and Social Overtures. The two factors were allowed to freely correlate. The comparison model was a one-factor model in which all six indicators were set to load on the same factor. In both models, year group was included as a covariate, regressing each factor onto year group to control for age differences in factor scores. We used full information maximum likelihood estimation to account for missing data, and computed robust standard errors in case of data not being normally distributed. A chi-square test with Satorra-Bentler correction was used to evaluate whether the two-factor model fitted better. We report confirmatory fit indices (CFIs) and root mean square error of estimation (RMSEA) with 90% confidence intervals. R package SemPlot (Epskamp & Stuber, 2017) was used to make a visualisation of the factor model.

Results

Reliability Analysis

First, we assessed whether each of the tests reliably tapped an underlying cognitive ability.

We supply classical test theory coefficients and item-level statistics in Tables 2 and 3.

Table 2

Reliability Analysis, including Cronbach's alpha and 95% confidence limits, standard error of measurement (SEm), and Revelle's beta. We also provide RMSEA when running IRT models, which gives an indication of unidimensionality of the test.

Test	Alpha	Alpha 95% CI	SEm	Beta	IRT RMSEA
Implicature Comprehension Test	.79	.76, .82	2.22	.68	.040
Children's Test of Local Textual Inference	.71	.66, .75	1.44	.60	.028
Pragmatic Violations	.73	.69, .77	1.57	.48	.073
Social Overtures	.82	.79, .85	1.51	.59	.086
Receptive Vocabulary	.79	.76, .82	1.99	.63	.024
Receptive Grammar	.89	.87, .91	2.05	.79	.059
Animal Matrices	.81	.78, .84	1.60	.64	.031

Table 3

Item-level statistics for each test: corrected item-total correlations (totals excluding the item) and item-level accuracy.

Test	Item-total correlations			Item-level Accuracy		
	Lower quartile	Median	Upper quartile	Lower quartile	Median	Upper quartile
Implicature Comprehension Test	.25	.29	.32	.69	.77	.84
Children's Test of Local Textual Inference	.20	.29	.33	.77	.88	.92
Pragmatic Violations	.24	.32	.35	.52	.69	.78
Social Overtures	.27	.36	.47	.81	.85	.89
Receptive Vocabulary	.17	.25	.31	.57	.89	.96
Receptive Grammar	.27	.35	.45	.80	.89	.92
Animal Matrices	.36	.40	.47	.37	.56	.71

Pragmatic and core language skills

As can be seen here, the tests show reasonably high internal consistency, so scores are subject to relatively little measurement error. The beta reliability coefficients are somewhat lower than the alpha coefficients. As the beta coefficient (worst split half reliability) reflects general factor saturation more accurately than the alpha coefficient (average split half), the discrepancies observed between the two types of reliability coefficients suggest that items in individual tests are not homogeneous. This is because item-total correlations are not universally high – this is likely due to item-level accuracy being variable, and we would not, for instance, expect to find strong correlations between total scores and accuracy on items with ceiling effects. The IRT RMSEA values indicate that a unidimensional model fitted each test at least adequately according to the criteria of .05 for ‘close fit’ and .089 for ‘adequate fit’ set out by Maydeu-Olivares and Joe (2014).

This reliability analysis was based on final versions of the tests. A few items were dropped from the tests because they were not reliable. They had flat ICCs (corresponding to chance-level accuracy and low item-total correlations) and so did not reliably tap a latent ability across the age-span. The weak items included one overtures item, four implicatures items, one vocabulary item, and two matrices items.

We assessed the extent to which our tests measured a stable ability by calculating test-retest correlations across two administrations two weeks apart. Correlations were as follows: Receptive Vocabulary .77 ($n = 34$), Animal Matrices .88 ($n = 26$), Implicature Comprehension .77 ($n = 77$), and Pragmatic Violations .52 ($n = 55$). The tests all showed reasonable test-retest reliability with the exception of the Pragmatic Violations.

As the tests were novel, it was important to assess their convergent validity with existing standardised measures, so a small subset of children aged 9 to 10 ($n = 20$) were given two WASI sub-tests. Correlation between the novel vocabulary task and WASI Vocabulary was .69, and between the Animal Matrices and WASI Matrix Reasoning was .70.

Hypothesis-testing

See Table 4 for descriptive statistics for each measure, and Table 5 for a correlation matrix. The Appendix presents a matrix of scatterplots showing relationships between the variables. The tests showed small to moderate correlations.

Table 4

*Descriptive Statistics. Variables included in the factor analysis are in **bold** type.*

	N	Mean	SD	Min	Max	Skew	Kurtosis
Implicature Comprehension Test	330	25.20	4.85	8	33	-0.68	0.18
Control items for Implicature	328	8.68	1.76	2	10	-1.60	2.13
Children's Test of Local Textual Inference	294	16.40	2.66	6	20	-1.32	2.17
Pragmatic Violations	320	11.90	3.03	0.48	18	-0.95	1.35
Social Overtures	297	19.60	3.57	8	23	-1.18	0.56
Receptive Vocabulary	338	29.60	4.35	10	39	-0.58	0.78
Receptive Grammar	291	41.40	6.17	14	49	-1.43	2.00
Animal Matrices	313	8.86	3.68	0	16	-0.31	-0.63

Pragmatic and core language skills

Table 5

Correlations between variables using pairwise complete observations.

	Textual Inference	Pragmatic Violations	Social Overtures	Vocabulary	Grammar	Matrices	Year Group
Implicature Comprehension Test	.34	.29	.42	.32	.36	.36	.23
Children's Test of Local Textual Inference		.11	.29	.36	.31	.41	.25
Pragmatic Violations			.33	.19	.36	.26	.14
Social Overtures				.43	.53	.39	.39
Receptive Vocabulary					.56	.44	.47
Receptive Grammar						.44	.54
Animal Matrices							.46

Pragmatic and core language skills

Our first hypothesis was that implicature comprehension would be relatively distinct from grammar and vocabulary skills. For this hypothesis, we ran a hierarchical multiple regression. The first stage of the regression, with year group as a predictor, was significant, $F(1, 223) = 5.73$, $p = .017$, and explained 2% of the variance in implicature scores on the Implicature Comprehension Test. Incorporating Receptive Vocabulary and Receptive Grammar as predictors significantly improved the model in the second stage, $F(2, 221) = 14.06$, $p < .001$, and predicted 12% of the variance in implicature scores, bootstrapped CI [6%, 23%]. In the third stage, narrative-based inferencing, as measured by the Children's Test of Local Textual Inference, also improved the model, $F(1, 207) = 14.25$. See Table 6 for the significance of individual predictors at each stage of the regression. The final model was significant, $F(4, 207) = 11.29$, and explained 16.3% of the variance, bootstrapped 95% CI [5.4%, 24.4%]. Compared to the sample of adults reported in Wilson and Bishop (2019), this is a similar proportion of variance; in adults, comparable tests of vocabulary, grammar and inferencing predicted 11.1% of variance in implicature comprehension, bootstrapped 95% CI [0%, 21.4%]. The results supported the hypothesis that implicature comprehension was only related to core language skills to a limited degree.

Table 6

Coefficients for Multiple Regression, with Implicature scores as the criterion variable.

	Estimate	SE	t-value	p-value
Stage 1				
Year group	0.54	0.30	1.77	.079
Stage 2				
Year group	0.00	0.31	-0.02	.984
Receptive Vocabulary	0.15	0.14	1.06	.289
Receptive Grammar	0.08	0.02	3.76	<.001
Stage 3				
Year group	-0.11	0.30	-0.36	.719
Receptive Vocabulary	0.06	0.14	0.41	.680
Receptive Grammar	0.06	0.02	3.28	.001
Children's Test of Local Textual Inference	0.14	0.04	3.78	<.001

In the second hypothesis, we stated that implicature comprehension would cluster with other tests that theoretically involve pragmatic processing but would be relatively distinct from core language skills. To test this hypothesis, we compared a two-factor model of language processing with a one-factor model. The two-factor model fitted the data well, CFI = .97, RMSEA = .06, 90% CI [.03, .09] whereas the one-factor model was weaker, CFI = .94, RMSEA = .08, 90% CI [.05, .10]. The difference in fit between the models was significant, $\chi^2(2) = 20.05$, $p < .001$. In the two-factor model, the correlation between the “core language” and “pragmatic language” factors, when controlling for the effect of age on each factor, was .79, 95% CI [.64, .93]. This indicates that, while a two-factor model does show a better fit, the factors are highly correlated in typically developing children. See Figure 1 for a visual representation of the factor model, and see Figure 2 for a plot of “core language” against “pragmatic language” factor scores extracted from the model. This plot shows a cloud of points positively correlated, with no

Pragmatic and core language skills

obvious outlying individuals – i.e. there are no clear cases of children having well-developed core language but weaker pragmatics, or vice versa.

Follow-up analyses

We assessed the relationship between nonverbal reasoning and the two language factors, and also tested for sex differences in the factors. For both these models, we ran the same two-factor model as shown in Figure 1, but added one additional covariate with regression paths to both factors. In the first analysis, this additional covariate was nonverbal reasoning, measured by the Animal Matrices. This model showed good fit, CFI = .97, RMSEA = .06, 90% CI [.03, .08]. This showed moderate relationships with “pragmatic language”, .50 95% CI [.38, .62], and “core language”, .39 95% CI [.27, .52]. The inclusion of nonverbal reasoning as a covariate had very little effect on the correlation between the “core language” and “pragmatic language” factors, which was now .74, 95% CI [.58, .90], indicating that the factors were strongly related due to their linguistic content rather than because they tapped general cognitive ability.

Next, Sex replaced nonverbal ability as a covariate. This model showed good fit, CFI = .96, RMSEA = .06, 90% CI [.03, .09]. Sex showed small effects on “pragmatic language”, .20 95% CI [.08, .32], and “core language”, .15 95% [.04, .25], with girls slightly outperforming boys on both factors when controlling for year group (i.e. age). In this model, year group retained large effects of age on “pragmatic language”, .56 95% CI [.46, .66], and “core language”, .71 95% [.63, .80].

All factor models were run with (i) all data and only complete cases, and (ii) non-transformed and transformed data, to evaluate the sensitivity of the results to the data processing.

Pragmatic and core language skills

Without transformations, there were several multivariate outliers, and with transformations, there were none. All factor analysis results were very similar regardless of processing. Despite the fact that two-factor models were preferred, it should be noted that the pattern of zero-order correlations between language tests shown in Table 5 does not clearly support a two-factor model, i.e. the pragmatic tests do not show a tight cluster separate from the core language tests. Therefore, exploratory factor analysis (implemented with R package psych; Revelle, 2017) was used to assess how well a data-driven analysis corresponded with the confirmatory factor analysis. First of all, the correlation matrix of all variables in the two-factor model was decomposed into eigenvalues to determine the number of factors to extract. Eigenvalues represent the proportion of variance accounted for by factors, in relation to the variance of individual variables being factor analysed, with eigenvalues below one accounting for less variance than an individual variable. Only one eigenvalue was above one (3.12), and so according to the Kaiser criterion only one factor was extracted. This factor accounted for 36% of variance, and represented core language skills, as can be seen in the large factor loadings for Receptive Vocabulary and Receptive Grammar shown in Table 7. Given that we only had one eigenvalue above one, we can conclude that the pragmatic tests did not naturally group together as a second factor. While this result questions the two-factor model tested by confirmatory factor analysis, it should also be noted that this exploratory factor analysis did not provide clear evidence for a one-factor model either, since the pragmatic tests (except the Social Overtures) did not show high factor loadings. This indicates that the pragmatic tests measured skills that were somewhat separate from core language, suggesting that language processing is multifactorial.

Table 7

Factor loadings for an exploratory factor analysis of the language test battery.

Test	Factor loading
Implicature Comprehension Test	.53
Children's Text of Local Textual Inference	.45
Pragmatic Violations	.39
Social Overtures	.69
Receptive Vocabulary	.71
Receptive Grammar	.77
Year group	.59

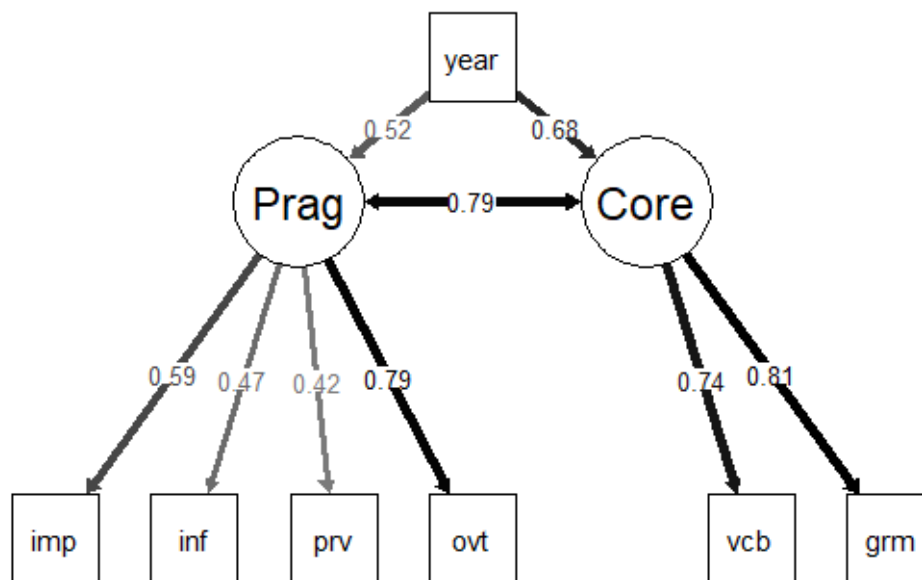


Figure 1. A two-factor correlated-traits model including a “pragmatic language” factor (Prag) and a “core language” factor (Core). Year group (year) has been included as a covariate to control for age effects in the language factors. imp = Implicature Comprehension Test; inf = Children’s Test of Local Textual Inference; prv = Pragmatic Violations; ovt = Social Overtures; vcb = Receptive Vocabulary; grm = Receptive Grammar.

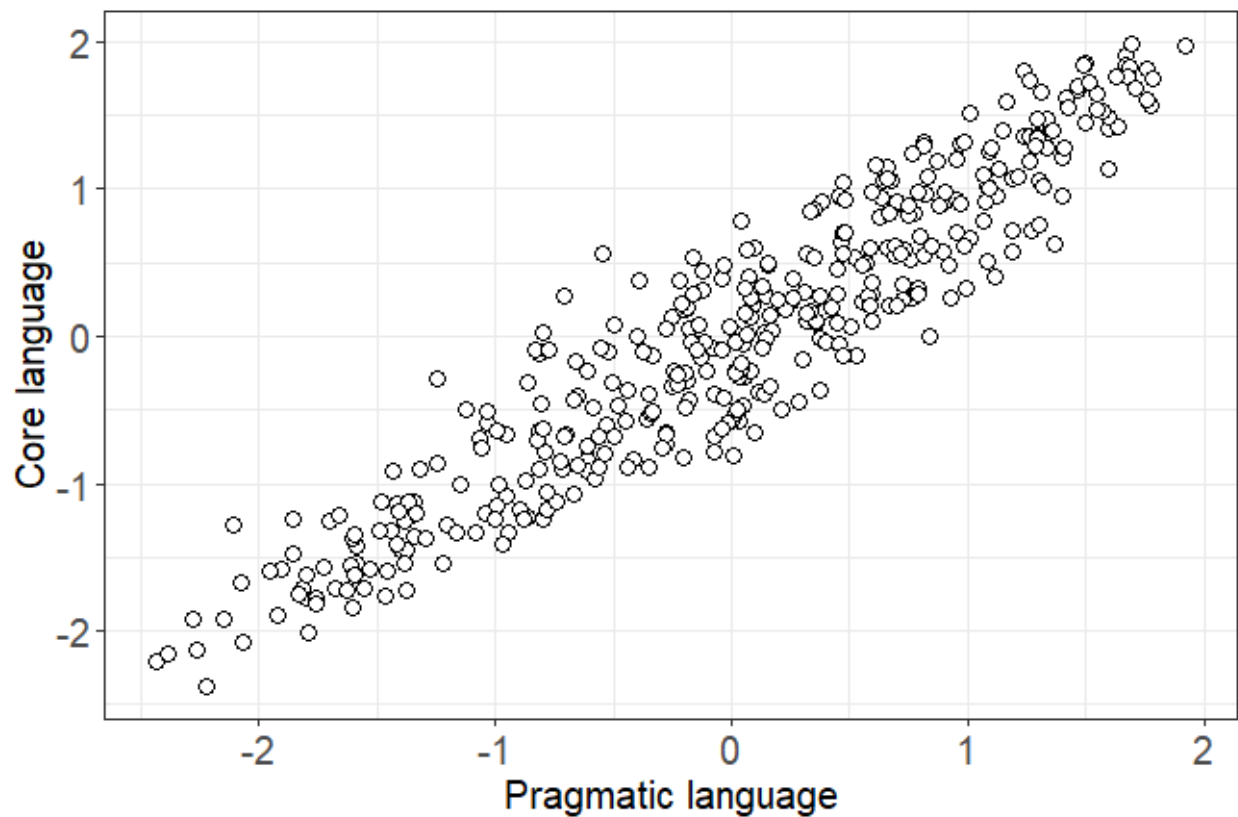


Figure 2. Plot showing “pragmatic language” factor scores against “core language” factor scores. Scores have been z-transformed, so the mean for each factor is zero, and each unit represents a standard deviation.

Discussion

Our analysis supported a multifactorial view of language processing, as a one-factor model of language comprehension did not give a strong fit to data collected in almost 400 children aged 7 to 13 years. Instead, confirmatory factor analysis gave greater support to a two-factor model incorporating separable, but highly correlated, factors for core and pragmatic aspects of language comprehension. The close relationship between these different aspects of language is in line with previous research (Matthews et al., 2018), even though the tests used here

Pragmatic and core language skills

set out to measure pragmatics more distinctly than has generally been attempted before. The findings support the view that well-developed core language abilities support development of those in the pragmatic domain, and vice versa. However, exploratory factor analysis did sound a note of caution about the coherence of the pragmatic “domain”. While this follow-up analysis provided further evidence that the skills measured by the pragmatic tests were somewhat separable from core language, it also showed that the pragmatic tests did not cluster together very well. This indicated that language processing might be best understood as multifactorial, with a coherent core language domain alongside multiple, somewhat separate pragmatic skills. The theoretical and practical implications of our findings run in perhaps rather different directions, so we discuss these separately below.

Theoretical Implications

As it was possible to measure core and pragmatic language as separable (if highly related), the findings support linguistic theories that language comprehension is not unitary but depends on separable processes of (i) decoding words and grammatical structures based on stored knowledge and (ii) inferring intended meanings in a flexible, context-dependent way (Cummings, 2007; Sperber & Wilson, 1986). Linguistic theories of pragmatics typically agree on a broad distinction between decoding and inference when contrasting semantics (“dictionary meaning” for want of a better term) and pragmatics (context-dependent meaning; Ariel, 2010). While the confirmatory factor analysis does support the view that there are somewhat separate cognitive processes underpinning language comprehension, this should not be overstated as there is substantial overlap. Interestingly, this overlap is not accounted for by general cognitive ability. As part of the analysis, we included nonverbal reasoning in the model as a covariate. On the one

Pragmatic and core language skills

hand, it was moderately associated with both language factors, as would be expected based on the idea that the g-factor (or general intelligence) affects all cognitive tests (Jensen, 1998). However, when controlling for nonverbal reasoning ability, the relationship between core and pragmatic aspects of language remained strong, suggesting that this relationship was due to the linguistic nature of the tests rather than because they are simply reflections of general cognitive ability.

As noted above, we should be cautious about the coherence of a pragmatic “domain”. The pragmatic tests showed fairly small correlations, suggesting that they made quite task-specific demands, rather than relying heavily on a domain-level ability. These low correlations could reflect something about the tests (perhaps, participants tended to adopt specific strategies that did not transfer between tasks) or something about the pragmatic domain itself (that it is less a “domain” and more a set of skills). While we cannot differentiate between these possibilities, it is notable that theoretical approaches tend to view pragmatics as heterogeneous, as noted by Matthews et al. (2018, p. 186): “While the domain of pragmatics is not clearly delineated or easily defined (Ariel, 2010), the family of pragmatic skills traditionally includes the ability to initiate conversation, to respond with contingent, relevant, and new information, to produce and understand utterances by drawing on context (including the perspectives of interlocutors and what is in their common ground), to use an appropriate register (respecting social status), to recount cohesive and coherent narratives and to understand non-literal language including irony.” Our data are quite consistent with the notion of a “family of pragmatic skills”.

There is a view that some individuals may have pragmatic difficulties with implied meaning in the absence of a core language impairment. For instance, the communication profile of verbal autistic individuals may involve well-developed vocabulary/grammar but differences in

ability and/or tendency to make inferences (Loukusa & Moilanen, 2009; Wilson & Bishop, 2020) and a preference for unusually literal language (Hobson, 2012). Based on these observations, we might hypothesise that processing implicature involves more than just core language skills, and our analysis supports that view. In the regression analysis, vocabulary and grammatical skills were only a modest predictor children's implicature scores. It is not surprising that these skills should play some role, as we must understand the "explicit content" of an utterance to have any chance of picking up an implied meaning, but clearly vocabulary and grammatical skills are not sufficient by themselves. It is notable that the Children's Test of Local Textual Inference was also a predictor of children's implicature scores, but again only explained a modest amount of variance. Narrative-based inferencing requires individuals to make inferences across a story to fill in "blanks" (Garnham & Oakhill, 1992). We expected to find some overlap between scores on this measure and the implicature comprehension test, since both require the integration of information across more than one sentence. However, the modest amount of overlap suggests that making inferences across narrative is quite different to picking up implied meanings in conversation, where we need to detect whether an implied response might have been intended.

Practical Implications

There is the longstanding intuition that core and pragmatic abilities may be dissociable, owing to clinical observation of individuals with communication difficulties that apparently affect pragmatics independently of core language; such individuals might be labelled as having pragmatic language impairment (PLI; Bishop, 1998). While this label was not formalised in any diagnostic manual, DSM-5 has incorporated a new diagnosis which is not dissimilar to PLI: social (pragmatic) communication disorder (American Psychiatric Association, 2013). This

Pragmatic and core language skills

requires social communication difficulties in the absence of both (i) structural language problems and (ii) repetitive and restrictive behaviours and interests characteristic of autism. However, this diagnosis has been the subject of controversy, not least because researchers question the validity of an isolated pragmatic impairment (see Norbury, 2014; Swineford, Thurm, Baird, Wetherby, & Swedo, 2014 for reviews). For instance, attempts to differentiate children with pragmatic language impairments from those with core language impairments on the basis of core language have not been very successful (Botting & Conti-Ramsden, 2003; Gibson, Adams, Lockton, & Green, 2013). Similarly, children said to have core language impairments are frequently found to have social communication impairments too upon standardised assessment (Leyfer, Tager-Flusberg, Dowd, Tomblin, & Folstein, 2008).

Our data speak to these issues in two ways. On the one hand, the confirmatory factor analysis supported the finding that pragmatic and core language impairments often co-occur, as the language factors were highly correlated. Looking at Figure 3, which plots “pragmatic language” and “core language” factor scores against each other, we see that some individuals scored considerably below their peers, but this was consistent across both factors. It is difficult to pick out any child with a specific pragmatic difficulty on visual inspection. To take a more formal approach, we could operationalise criteria as a “pragmatic language” factor score at least one SD below the mean and a discrepancy of at least one SD between “pragmatic language” and “core language” factor scores. No child showed this profile of a specific pragmatic difficulty. Of course, this might be a feature of the sample recruited here. Our sample was recruited from mainstream school, and if children with specific pragmatic problems mostly attend special educational settings, then our sample would have excluded them. Therefore, our results

Pragmatic and core language skills

demonstrate that core and pragmatic language abilities are closely associated in typical development, and if some individuals in clinical samples show a discrepancy between skills, this is unusual and not part of typical variation.

Our analysis also highlights the potential importance of making a distinction between skills and domains. The confirmatory factor analysis described here aimed to measure broad domains of pragmatics and core language by drawing upon what is common across multiple tests, but as the exploratory factor analysis indicated, we should perhaps question the existence of a pragmatic “domain”, which might be best understood as a “family of skills”. It is, therefore, entirely plausible that some children may show a spiky profile, in which some test-specific skills are stronger than others. In this regard, we might reflect on the relatively low relationship ($r = .34$) observed between the Implicature Comprehension Test and Children’s Test of Local Textual Inference. Both tests require the child to draw inferences, in short conversational interchanges in the former and short narratives in the latter, but the modest correlation indicates that there are somewhat different skills involved, and some children might find one task more challenging than the other. It is important, therefore, that where educationalists and clinicians are assessing children’s ability with implied and inferred meanings that they assess inferencing in a range of situations, including conversation and narrative, as inferencing is multifaceted. Overall, the message might be that domain-level ability in core language is likely to predict how a child will perform across a range of different language tests, but they might show some variability in task-specific skills relevant to pragmatic aspects of communication.

We should bear in mind a couple of limitations of this study. Firstly, it was not feasible to incorporate a screening questionnaire for language impairment, and this means we cannot assess

the level of agreement between our tests and informant report in identifying possible language impairments. In particular, there is no way of knowing whether the tests were effective in capturing the difficulties of children with specific problems with pragmatic aspects of communication. Traditional language tests fail to capture these difficulties (Conti-Ramsden, Crutchley & Botting, 1997), and while pragmatic assessment was a special focus in this study, it is possible that our ability to detect pragmatic difficulties may have been limited by the measures used here. The nature of pragmatics is flexibility – the ability to adapt our use and understanding of language to the social context – and this is inherently difficult to operationalise in a structured testing situation.

We present evidence for a multifactorial model of language processing in typically developing children. Core language represents a coherent language domain involving vocabulary and grammatical skills, whereas pragmatic language abilities seem to represent a family of skills that are more heterogeneous and somewhat separable from core language.

References

- Altman, N., & Krzywinski, M. (2016). Regression diagnostics. *Nature Methods*, 13(5), 385–386. doi:[10.1038/nmeth.3854](https://doi.org/10.1038/nmeth.3854)
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5 (5th edition)*. Arlington, VA: American Psychiatric Publishing, Inc.
- 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:[10.3389/fpsyg.2017.00996](https://doi.org/10.3389/fpsyg.2017.00996)

- Antoniou, K., & Katsos, N. (2017). The effect of childhood multilingualism and bilingualism on implicature understanding. *Applied Psycholinguistics*, 38, 787–833.
- Ariel, M. (2010). *Defining pragmatics*. Cambridge, UK: Cambridge University Press.
- Bishop, D. V. M. (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:[10.1017/S0021963098002832](https://doi.org/10.1017/S0021963098002832)
- Bishop, D. V. M., & Adams, C. (1989). Conversational characteristics of children with semantic-pragmatic disorder. II : What features lead to a judgement of inappropriacy ? *British Journal of Disorders of Communication*, 24(3), 241–263. doi:[10.3109/13682828909019890](https://doi.org/10.3109/13682828909019890)
- Botting, N., & Conti-Ramsden, G. (2003). Autism, primary pragmatic difficulties, and specific language impairment: Can we distinguish them using psycholinguistic markers? *Developmental Medicine and Child Neurology*, 45(8), 515–524. doi:[10.1017/S0012162203000963](https://doi.org/10.1017/S0012162203000963)
- Canty, A., & Ripley, B. (2017). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-19.
- Carston, R. (2002). Linguistic meaning, communicated meaning and cognitive pragmatics. *Mind & Language*, 17(1-2), 127–148. doi:[10.1111/1468-0017.00192](https://doi.org/10.1111/1468-0017.00192)
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29. doi:[10.18637/jss.v048.i06](https://doi.org/10.18637/jss.v048.i06)

- Conti-Ramsden, G., Crutchley, A., & Botting, N. (1997). The extent to which psychometric differentiate subgroups of children with SLI. *Journal of Speech, Language, and Hearing Research*, 40(4), 765–77. doi:[10.1044/jslhr.4004.765](https://doi.org/10.1044/jslhr.4004.765)
- Cummings, L. (2007). Clinical pragmatics: A field in search of phenomena? *Language and Communication*, 27(4), 396–432. doi:[10.1016/j.langcom.2007.06.001](https://doi.org/10.1016/j.langcom.2007.06.001)
- Devine, R. T., & Hughes, C. (2016). Measuring theory of mind across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *Journal of Experimental Child Psychology*, 149, 23–40. doi:[10.1016/j.jecp.2015.07.011](https://doi.org/10.1016/j.jecp.2015.07.011)
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Epskamp, S., & Stuber, S. (2017). semPlot: Path diagrams and visual analysis of various SEM packages' output. R package version 1.1. Retrieved from <https://cran.r-project.org/package=semPlot>
- Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Quantitative methods in education and the behavioral sciences: Issues, research, and teaching. structural equation modeling: A second course* (pp. 439–492). Charlotte, NC: Information Age Publishing.
- Flora, D. B., Labrish, C., & Chalmers, R. P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology*, 3(55). doi:[10.3389/fpsyg.2012.00055](https://doi.org/10.3389/fpsyg.2012.00055)

Garnham, A., & Oakhill, J. (1992). Discourse processing and text representation from a “mental models” perspective. *Language and Cognitive Processes*, 7(3), 193–204.

doi:[10.1080/01690969208409384](https://doi.org/10.1080/01690969208409384)

Gibson, J., Adams, C., Lockton, E., & Green, J. (2013). Social communication disorder outside autism? A diagnostic classification approach to delineating pragmatic language impairment, high functioning autism and specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 54(11), 1186–1197.

doi:[10.1111/jcpp.12079](https://doi.org/10.1111/jcpp.12079)

Hobson, R. P. (2012). Autism, literal language and Concrete thinking: Some developmental considerations, 27, 4–21. doi:[10.1080/10926488.2012.638814](https://doi.org/10.1080/10926488.2012.638814)

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jorgensen, T. D., Pornprasertmanit, S., Miller, P., & Schoemann, A. (2018). simsem: Simulated structural equation modeling. R package version 0.5-14. Retrieved from <https://cran.r-project.org/package=simsem>

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A., Rosseel, Y., Miller, P., Quick, C., ... Preacher, Kristopher Coffman, D. (2019). semTools: Useful tools for structural equation modeling. R package version 0.5-2. Retrieved from <https://github.com/simsem/semTools/wiki>

Kalandadze, T., Norbury, C. F., Nærland, T., & Næss, K.-a. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review.

doi:[10.1177/1362361316668652](https://doi.org/10.1177/1362361316668652)

Pragmatic and core language skills

- Korkmaz S, Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2), 151–162.
- Leyfer, O. T., Tager-flusberg, H., Dowd, M., Tomblin, J. B., & Folstein, S. E. (2008). Overlap between autism and specific language impairment: Comparison of Autism Diagnostic Interview and Autism Diagnostic Observation Schedule scores, (November), 284–296. doi:[10.1002/aur.43](https://doi.org/10.1002/aur.43)
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule - Second Edition (ADOS-2)*. Torrance, CA: Western Psychological Services.
- Loukusa, S., & Moilanen, I. (2009). Pragmatic inference abilities in individuals with Asperger syndrome or high-functioning autism. A review. *Research in Autism Spectrum Disorders*, 3(4), 890–904. doi:[10.1016/j.rasd.2009.05.002](https://doi.org/10.1016/j.rasd.2009.05.002)
- Mangiafico, S. (2018). rcompanion: Functions to support extension education program evaluation. R package version 1.13.2. Retrieved from <https://cran.r-project.org/package=rcompanion>
- 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:[10.1080/15475441.2018.1455584](https://doi.org/10.1080/15475441.2018.1455584)
- Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49(4), 305–328. doi:[10.1080/00273171.2014.911075](https://doi.org/10.1080/00273171.2014.911075)

- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3), 381–401. doi:[10.1016/j.jml.2006.06.006](https://doi.org/10.1016/j.jml.2006.06.006)
- Meinhardt-Injac, B., Daum, M. M., Meinhardt, G., & Persike, M. (2018). The two-systems account of theory of mind: Testing the links to social-perceptual and cognitive abilities. *Frontiers in Human Neuroscience*, 12(25). doi:[10.3389/fnhum.2018.00025](https://doi.org/10.3389/fnhum.2018.00025)
- Norbury, C. F. (2014). Practitioner Review: Social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 55(3), 204–216. doi:[10.1111/jcpp.12154](https://doi.org/10.1111/jcpp.12154)
- R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Revelle, W. (2017). psych: Procedures for personality and psychological research. R package version 1.7.8. Retrieved from <https://cran.r-project.org/package=psych>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. Retrieved from <http://www.jstatsoft.org/v48/i02/>.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford, England: Blackwell.
- Stevenson, C. E., Heiser, W. J., & Resing, W. C. M. (2016). Dynamic testing of analogical reasoning in 5- to 6-year-olds: Multiple-choice versus constructed-response training items.

Journal of Psychoeducational Assessment, 34(6), 550–565.

doi:[10.1177/0734282915622912](https://doi.org/10.1177/0734282915622912)

Swineford, L. B., Thurm, A., Baird, G., Wetherby, A. M., & Swedo, S. (2014). Social (pragmatic) communication disorder: A research review of this new DSM-5 diagnostic category. *Journal of Neurodevelopmental Disorders*, 6(41). doi:[10.1186/1866-1955-6-41](https://doi.org/10.1186/1866-1955-6-41)

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Wilson, A. C., & Bishop, D. V. M. (2019). "If you catch my drift...": Ability to infer implied meaning is distinct from vocabulary and grammar skills. *Wellcome Open Research*, 4(68). <https://doi.org/10.12688/wellcomeopenres.15210.2>

Wilson, A. C., & Bishop, D. V. M. (In publication). "Second guessing yourself all the time about what they really mean...": Cognitive differences between autistic and non-autistic adults in understanding implied meaning. *Autism Research*.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133. doi:[10.1007/s11336-003-0974-7](https://doi.org/10.1007/s11336-003-0974-7)